

New research council needed

Ad hoc collaborations between Europe's national funding agencies need to be complemented and, to some extent, replaced by a funding agency charged with supporting outstanding researchers from across the continent.

"There is nothing more difficult and dangerous, or more doubtful of success, than an attempt to introduce a new order of things in any state. For the innovator has for enemies all those who derived advantages from the old order of things, whilst those who expect to be benefited by the new institutions will be but lukewarm defenders. This indifference arises in part from fear of their adversaries who were favoured by the existing laws, and partly from the incredulity of men who have no faith in anything new that is not the result of well-established experience."

This sixteenth-century quote by Niccolò Machiavelli is one that Wilhelm Krull, secretary-general of Germany's Volkswagen Foundation, likens to the task of those like him who share a bold vision of a European Research Council (ERC) that is independent of the European Commission and national research bodies, and has a mandate to fund basic and strategic research anywhere in Europe on the sole basis of scientific excellence and innovation.

Krull's assessment would have been echoed by Pierre Werner, a European visionary who died last month but who in 1960 fathered the concept of the euro. The birth of the euro is relevant because, despite the many questions surrounding the *raison d'être* of an ERC, creating one is a picnic compared to having persuaded the currency's 12 member states to abandon their beloved deutschmarks, francs, lire and pesetas to adopt a single currency and hand over control of economic policy to a central European Bank.

The euro succeeded because of sustained political commitment at the highest level and unprecedented cooperation among Europe's banks. If an effective ERC is to see the light of day, the same is now needed from national politicians and research agencies.

Why do we need an ERC? In short, because it can help Europe make better use of its scientific resources. A European-wide funding scheme for selected categories of basic research would cause the cream to rise and would focus spending better. It could also help fledgling and unfashionable disciplines to obtain decent levels of support in a few well-funded labs.

Moreover, in areas such as bioinformatics and systems biology there is a shift towards truly multidisciplinary approaches, cutting across departmental boundaries and embracing partnership, rather than competition, between major research universities. Such grass-roots network initiatives are growing up all over Europe.

Best practice

Bringing these initiatives under one roof should rationalize peer review and other procedures. A study of existing best practices of peer review and evaluation should be essential in planning an ERC, and an invaluable exercise in itself from which Europe's various research bodies and networks might learn. It would also facilitate international cooperation — single points of contact, instead of multiple negotiations with national bodies.

The devil is in the details, and several bodies have set up task forces to explore these. A conference in October hosted by the Danish research ministry will gather Europe's scientific elite. Its agenda asks the right questions: do we need an ERC, what should it do, who should pay for it, who should run it, and how do we get there?

There are some fundamental conditions that need to apply. Whoever funds it, an ERC must be politically independent. Critically, it must also avoid the bureaucracy that has characterized the European Union's joint Framework research programmes, and demonstrate an impeccable commitment to scientific excellence.

The ERC's day-to-day operations should be run by scientists for scientists, and their choices used to irrigate Europe's existing research powerhouses rather than compete with them. But it needs political legitimacy and clout. Approval of a strategic vision would be needed from Europe's research ministers.

This arrangement may also help to resolve the question of whether the ERC should be bottom-up, supporting largely investigator-driven proposals, or strategically driven. Most current proposals favour a bottom-up approach, but the ERC will, at least initially, have limited funds, and some strategic oversight of research priorities will be needed if money is not to be spread too thinly. Risky areas of research by outstanding scientists could be one priority, transdisciplinarity another. Decisions will also need to be made as to research areas that would have a clear added value by being managed at a European level, as opposed to through the coordination of existing national efforts.

Funding plan

Who should fund an ERC? One proposal is that a levy of 5% could be made on national research organizations. But European Union money should also be considered. Indeed, a case could be made for redirecting some Framework funding, which currently targets economic competitiveness rather than basic research. Charities and non-governmental organizations might find it worth buying into as well.

Prototypes already exist. The European Science Foundation last year created EUROCORES, which, after a call for proposals and peer review, recommends the best projects for funding. It is then up to national members to decide whether they want to provide funding or not. This has the attraction of avoiding pooling funds — a recipe for political discord, as the research projects supported may not tally with nations' own interests. A similar funding variant is that of the European Space Agency, in which members pay a fixed subscription to a compulsory science programme. But there is also an additional 'optional programme', in which member states can decide whether or not to support a particular project.

Some form of variable geometry seems unavoidable in any ERC, as in practice it will be impossible to get a consensus on anything among the thousands of European research institutions. Even within member states, science agencies jealous of their power are still struggling to coordinate disciplines among themselves. Extrapolate that to 15 countries — and soon to 25 — and you see both the need for an ERC and the problems it faces.

If high-level support for the ERC can be sustained, it is not unrealistic to launch it from scratch as a full-blown agency. In one of several possible pilot schemes there could be several high-profile European research awards in the life sciences, in combination with a fellowship scheme supporting young investigators in setting up research groups in a country other than their own. A job for the European Molecular Biology Organization, perhaps? ■





No vote
German council
snubs European
neutron source
p262



Too few crew
Scaled-down space
station 'cannot do
high-priority science'
p263



Terror risk?
Polio experiment
leads to worries over
data restriction
p265



Angling alert
Maryland under
threat from invading
snakehead fish
p267

Government spending promise offers British research a boost

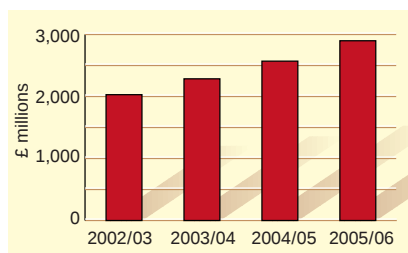
David Adam and Natasha McDowell, London

British researchers enjoyed some rare sunshine this week, as summer finally arrived and the government announced a £1.25-billion (US\$2-billion) increase in annual science funding over the next three years.

Gordon Brown, the Chancellor of the Exchequer, announced on 15 July that the science budget at the Department of Trade and Industry (DTI) — which supports most British university research — will grow from £2 billion in the current financial year to £2.9 billion in 2005–06. The hike includes an extra £400 million a year for research grants by the end of that period, as well as £100 million to raise salaries for young researchers. There will also be fresh funds to renovate university labs.

Researchers and lobby groups have welcomed the pledge, part of a £60-billion spending package in the public services. The spree is seen as an effort by the Labour government to commit to stronger public services, after five relatively austere years in office.

The science budget will grow by about 10% in real terms each year over the next three years (see chart), compared with 7% in the 1998 and 2000 reviews. The individual research councils that distribute the funds



Spending spree: Gordon Brown has pledged to increase funds for research supported by the DTI.

won't discover their shares until later this year.

"There is substantial new money and importantly much of it is aimed at solving some of the long-term problems suffered by the science research base," says Peter Cotgreave, director of the pressure group Save British Science. The increased funding remains merely a promise, but Cotgreave says he is confident that the government will deliver, "even under the most pessimistic predictions about the economy".

Within the Department of Trade and Industry's enlarged science budget, the research councils will also receive £120 million extra by 2005–06 to pay overheads for



research grants, as well as new funding to help researchers set up spin-off companies. And the Department for Education and Skills will provide an extra £200 million annually, by the end of the period, for university equipment and buildings.

Under the plan, Brown estimates that the average PhD grant will grow from £10,000 a year to £13,000, while the wage of a postdoctoral research fellow will increase by £4,000 to £21,000 a year.

The pledge did not cover research done in other government departments, such as the Ministry of Defence and the Department for Environment, Food and Rural Affairs.

Separately, the Wellcome Trust, Britain's largest biomedical charity, announced a new £280-million funding package over five years. The plan will include money to extend the Sanger Centre near Cambridge and to help the government build a national centre charged with improving science teaching. ■

California lab fires physicist over retracted finding

Rex Dalton, San Diego

The Lawrence Berkeley National Laboratory (LBNL) in California has fired a physicist for allegedly fabricating data that contributed to a 1999 article on the purported discovery of two heavy elements.

Victor Ninov was fired in May after a year of internal probes into claims of scientific misconduct involving the apparent discovery of elements 118 and 116, reported in *Physical Review Letters* (V. Ninov *et al. Phys. Rev. Lett.* 83, 1104; 1999). *Physical Review Letters* formally retracted the article on 15 July, although Ninov's co-authors withdrew their support a year ago.

Ninov, who could not be reached for comment, has filed a grievance with the

University of California, which manages the LBNL for the Department of Energy (DOE), says Pier Oddone, the LBNL's deputy director. Ninov, who was in charge of computer analysis of the experiments, denies fabricating any data, says Oddone.

None of Ninov's 14 co-authors on the article were engaged in data fabrication, says Oddone. The data in question involved a computer analysis of an experiment in which high-energy krypton ions were fired at a lead target in the LBNL's cyclotron.

But questions are being raised in Germany about Ninov's work in two earlier articles, published while he was working at the GSI, a national research institute in Darmstadt.

The *European Physical Journal A*

published an article this month (S. Hofmann *et al. Eur. Phys. J. A* 14, 147–157; 2002) discussing repeat experiments of the previously reported discoveries of elements 111 and 112 (S. Hofmann *et al. Zeitschrift für Physik A* 350, 277–280; 1995, and S. Hofmann *et al. Z. Phys. A* 354, 229–230; 1996). Ninov did analysis at the GSI for the 1995 and 1996 papers, researchers say.

The *European Physical Journal A* article alleges that there were two instances in which raw data from the earlier experiments did not match the published results. Results "were spuriously created", wrote GSI physicist Sigurd Hofmann, lead author in all three articles. But the discovery of elements 111 and 112 still stands, he wrote. ■

German council confounds plans for neutron project

Quirin Schiermeier, Munich

Plans for a major neutron source to be built in Europe suffered a significant setback last week, when Germany's science council declined to endorse the project.

The European Spallation Source (ESS) was one of nine proposed science projects assessed by the council in a report published on 12 July. But the neutron source received an unexpectedly harsh review, which has now raised doubts over whether Germany will provide its share of the funds needed for the project.

The council endorsed an assessment of the ESS by a subcommittee chaired by Hans Spiess, a director at the Max Planck Institute for Polymer Research in Mainz, saying that the demand for neutrons does not justify the estimated 1.4-billion-euro (US\$1.4-billion) total investment in an advanced neutron source. By 2011, the earliest date that the ESS could become operational, some of its studies could be done more cheaply using alternative technologies such as synchrotrons or electron microscopes, the council said in its statement.

Instead it argued that higher priority should be given to TESLA, an electron-positron collider proposed for construction at DESY, the high-energy physics research centre in Hamburg. TESLA, which is expected to cost 3.5 billion euros, will also feature a 670-million-euro X-ray free-electron laser, which will produce high-energy pulses that can be used to determine the structure of biological molecules, among other things (see *Nature* **415**, 110–111; 2002).

"The ESS would allow continuation of neutron research, if on a very high level," explains Spiess. "But TESLA is a truly visionary project that will open completely new ground."

The council's assessment has shocked European neutron scientists, who had hoped that construction of the ESS could start at a site in Europe by 2004. The facility would be used to probe the molecular structure of matter. Structural biologists, condensed-matter physicists and geologists, for example,



Bob Cywinski says Europe needs a new neutron source.

favour neutron sources because they can identify lighter atoms — such as hydrogen — in structures.

"It is nonsense to play off synchrotrons or free-electron lasers against neutrons. They are complementary leaps into the future," says Bob Cywinski, a physicist at the University of Leeds, and former chairman of the European Neutron Scattering Association. "If you take away Europe's leading edge in neutron research you take away our leadership in condensed-matter science as well," he adds.

Japan and the United States are already building neutron sources that are expected to come on line in 2006. With the ESS — which is a partnership between Germany, France, Britain and other European nations — researchers were hoping to retain their traditional leadership in neutron science (see *Nature* **417**, 883; 2002).

Two sites in Germany — the Research Centre Jülich in North Rhine Westphalia and a greenfield site in Saxony-Anhalt — have put in a bid to host the facility. But they need to redesign and streamline their applications, Spiess says, before the council will support them.

It is still possible that Germany will eventually host the ESS, officials say, or at least contribute to the project's costs if it is built elsewhere. But without a positive vote from its influential science council, the German government is unlikely to back the project.

At the very least, the council's verdict is expected to complicate informal negotiations between the countries involved in the project, which is still at an early stage. Britain, for example, has not yet even offered a candidate site for the neutron source.

Peter Tindemans, chairman of the ESS research council, said in a letter to Germany's science council, the Wissenschaftsrat, and to the science ministers of Saxony-Anhalt and North Rhine Westphalia that the German science council's assessment was not widely shared, and that a "European consensus view" would determine the project's future.

Tindemans says he thinks that the decision will delay the ESS, but not necessarily kill it. "The last word has not yet been spoken," he says, "but without a firm commitment from Germany, it will be extremely difficult to get the project off the ground." ■

Senate nod prompts fresh analysis of nuclear waste dump

Geoff Brumfiel, Washington

The US Senate's decision on 9 July to approve the proposed nuclear-waste repository at Yucca Mountain, Nevada, has set the stage for a fresh and exhaustive technical investigation of the plan.

The Department of Energy (DOE), which is running the project, must produce convincing data on hundreds of technical issues before starting to build the repository, in 2008 at the earliest.

The Nuclear Regulatory Commission (NRC), which must grant a licence before the project can begin, has a list of 293 separate topics that it wants answers on, officials there say. There remain "gaps in information" concerning the site's volcanic and seismic activity, hydrology and geology, as well as engineering issues, says Janet Schlueter, a branch chief at the NRC's Division of Waste Management.

Under the terms of the licence application, project managers have to satisfy the NRC that the repository will contain the waste safely for 10,000 years. And Schlueter stresses that the licensing process will not be a simple 'rubber stamp' — the NRC has a four-year window to interrogate the DOE over the proposal.

Energy undersecretary Robert Card says that the project team has collected data that will fulfil most of the NRC's requirements. "We've already resolved many of these issues," he says. The department hopes to file its application by the end of 2004.

But "there are several issues where there is significant work to be done," says George Hornberger, an environmental scientist at the University of Virginia and chairman of the NRC's Advisory Committee on Nuclear Waste. NRC officials have recently raised the possibility of volcanism at the site, says Hornberger. Both the DOE and the NRC are currently running computer simulations to address a disagreement about the risk posed by volcanic activity (see *Nature* **412**, 850–852; 2001).

Other questions surround the deterioration of the sophisticated metal containers that would hold the waste, says Kevin Crowley, director of the Board on Radioactive Waste Management at the National Research Council. Project planners "need to understand how that metal will behave over thousands of years," he says.

Hornberger is not convinced that the DOE will be able to complete its licence application by 2004. "They have a lot to do, and it's not going to be easy," he says. ■



Cool response: plans for a European neutron source have failed to win approval in Germany.

Cutbacks 'will cripple space station science'

Tony Reichhardt, Washington

The scaled-down International Space Station (ISS) currently proposed by NASA — with only three astronauts on board and very limited resources — would be able to do very little high-priority science, according to a review of research planned for it.

If the curtailed design is not enhanced, “NASA should cease to characterize the ISS as a science-driven programme”, charges the Research Maximization and Prioritization (REMAP) task force, which presented the outcome of its four-month assessment to NASA's Advisory Council on 10 July.

Startled members of the council warned that the assessment could lead to the abandonment of the lumbering, 18-year-old project. “If I were in the White House,” says council member Tom Young, a retired aerospace executive, “I would take this as a recommendation to terminate the existing space station.”

But most political observers say that it is too late for that, and that some version of the station will be completed. The White House budget office last year ordered NASA to plan for a three-member station crew, instead of a more capable six- or seven-person version, unless it can manage the project — currently at least \$5 billion over budget — more tightly (see *Nature* **410**, 399; 2001). The decision drew protests from scientists and from the station's international partners, including the European and Japanese space agencies.

Commissioned by NASA and chaired by Columbia University biopsychologist Rae Silver, REMAP was asked to set research priorities for the agency's Office of Biological and Physical Research (OBPR), which sponsors most of the station's planned experiments.

Surveying the broad portfolio of physics and biology research sponsored by the OBPR, the 20 members of REMAP gave “highest priority” ranking to more than a dozen subdisciplines, including studies of radiation health, crew behaviour and advanced life support. These were deemed either to have intrinsic scientific merit or to offer help for enabling future human space travel. Lower priority was given to experiments in protein-crystal growth, which have been criticized by other review groups (see *Nature* **404**, 114; 2000).

After determining the rankings, REMAP worked with the OBPR on an implementation plan — with discouraging results. By NASA's own estimates, a three-person crew could handle only a small fraction of the ‘high-priority’ research. Key resources, such as electrically powered ‘lockers’ for experiments, would be in critically short supply. Limiting the space shuttle to four flights a year, which has been proposed as a money-saving move, would leave almost no room



Up in the air: limiting the ISS's crew to three would leave onboard science struggling for resources.

for science equipment on station resupply missions. And two key pieces of lab hardware — a centrifuge for varying g-forces, and a holding facility for animals and plants — are in danger of being delayed or scrapped.

This has led REMAP to conclude that the scaled-down station would not be the lab that NASA originally envisioned. Some high-priority science could still be done on board, Silver says, but the station could no longer count research as its primary function.

Former astronaut and US senator John Glenn, a member of the advisory council, says

he is worried that the report will “be used as material to kill the whole programme”.

But Silver is hopeful that policy-makers will respond to the assessment by improving the specification of the station so that it can be used for valuable science. She says she is encouraged by the fact that NASA's administrator, Sean O'Keefe, has repeatedly encouraged her task force to identify the best science that could potentially be done on board the station, without worrying about the constraints that the scaled-down design would impose. ■

NASA aims to reach Pluto by 2020

Tony Reichhardt, Washington

Sending a spacecraft to Pluto and the distant Kuiper Belt should be the United States' top priority in Solar System exploration, according to a 'decadal survey' of planetary science — the most thorough attempt yet to set a long-term agenda for research in the field.

NASA already is heeding the advice, released on 11 July by the National Academy of Sciences, and hopes to fund the on-off New Horizons mission (see *Nature* **414**, 571; 2001) to reach Pluto by 2020, according to Colleen Hartman, who heads the agency's Solar System exploration office.

The review is meant to build on the successful tradition of decadal reports for astronomy and astrophysics, which have been influential in guiding US government funding (see *Nature* **405**, 381–382; 2000).

The proposed agenda for 2003 to 2013 calls for three classes of planetary missions.

Small (less than \$325 million) missions would essentially continue NASA's existing Discovery programme, which has sent spacecraft to study asteroids and comets. The medium-sized (up to \$650 million) category roughly matches the agency's New Frontiers line introduced this year, and would include the mission to Pluto and the Kuiper Belt, as well as a spacecraft to return samples from the Moon's south pole.

The most controversial recommendation is to revive the kind of large, expensive (more than \$650 million) flagship missions that NASA abandoned in the 1990s. The panel, chaired by planetary scientist Michael Belton of Belton Space Exploration Initiatives in Tucson, Arizona, picked a spacecraft to explore Jupiter's moon Europa as the first entry in this category. Belton says that NASA's projected budget should allow for one such project every ten years or so. But Hartman says that the money isn't there. ■

Publishers soldier on despite electronic bugs

Rex Dalton, San Diego

It would have made the perfect gift — were it not for the vagaries of electronic publishing. Attending the wedding of a former doctoral student, neuroscientist John Salamone from the University of Connecticut had hoped to confirm publication of a paper the pair had co-authored. Instead, he brought news of a problem that has left many other researchers equally frustrated — software used by the journal to keep track of the manuscript had misfired.

As scientific publishers react to the opportunities offered by the Internet and online publishing, electronic systems are pervading every aspect of the publication process. But with these advances often come teething troubles.

For Salamone, that meant a lengthy delay. The team submitted its paper to *Neuroscience* last November, but in April the journal's editors admitted that the manuscript-tracking software had lost both the manuscript and the reviewers' comments. "This was supposed to be a fast-track publication," says Brian Carlson, Salamone's former student who is now at the University of California, Los Angeles.

It took until June, a month after Carlson's wedding, for the paper to be accepted for publication. In the meantime, Carlson was turned down for a grant from a private foundation — which he attributes in part to the delayed publication — and a co-author had to hold back a second paper that couldn't be submitted until Carlson's was published.

The episode is fairly typical of the problems encountered as scientific journals move deeper into the electronic age, researchers say. Robert Dickinson, president of the American Geophysical Union (AGU), told the society's members in a statement earlier this month: "The most critical problem that has been facing AGU is its transition to electronic publishing."

A major issue for journals is keeping track of manuscripts. Because the publication process varies so much between journals, off-the-shelf software packages tend not to fit the bill, so scientific publishers often have to prepare or commission bespoke software for their titles. And, as with all customized software, this carries an element of risk.

Neuroscience, which is published by Dutch company Elsevier Science on behalf of the Paris-based International Brain Research Organization, had been testing a manuscript-tracking system called SMARTWorks, which was developed several years ago in San Diego. The system was written for Elsevier, and is unrelated to commercial software sold by SMARTworks.com of Dayton, Ohio.

David Amaral, a neuroscientist at the

University of California, Davis, and chief editor of *Neuroscience*, says that the system developed some "quite bizarre" problems, and the journal stopped using it in March. Elsevier officials say they have since reduced their reliance on SMARTWorks and are trying other systems.

Even when the software works, some scientific publishers have seen their software suppliers go bust, leaving them without any technical back-up. This happened to the *European Journal of Neuroscience*, published for the Federation of European Neuroscience Societies by the UK company Blackwell Science, when the software supplier it was



using, PaperPath of Massachusetts, closed its doors last September.

"It was working beautifully, absolutely perfectly," says Barry Everitt, a behavioural neuroscientist at the University of Cambridge and editor-in-chief of the bimonthly journal. "Then the company evaporated and our database went down." Everitt says his journal fell a month behind on its publishing schedule, after temporarily reverting to the use of paper manuscripts. "We recovered, but it was very damaging," he recalls.

The new systems are also raising concerns about security. In May, for example, a researcher at the National Institutes of Health told the Nature Publishing Group (NPG), publisher of *Nature*, that one of his postdocs had inadvertently accessed two papers by other authors while "playing around" with their own submission to *Nature Cell Biology*. The journal uses a manuscript system developed for NPG. Ruth Wilson, peer-review manager at NPG, says that security on the system has now been tightened so that the problem cannot recur.

Publishers are hoping researchers will be patient, and say that the transition to full-blown electronic publishing will eventually benefit everyone. "We are all running into problems as we scale up," says Jasna Markovac, a director of Elsevier based in San Diego. "But we will find solutions." ■

US panel split on research cloning

Kendall Powell, Washington

President George Bush's bioethics council has issued a split decision on whether to ban cloning for research purposes — increasing the likelihood that Congress will fail to legislate on cloning this year.

The President's Council on Bioethics issued a report on 11 July that unanimously endorsed a ban on human cloning to produce children. Ten of the council members recommended a four-year moratorium on cloning for research purposes, sometimes referred to as therapeutic cloning, whereas seven said that research cloning should proceed once appropriate regulations are in place.

Leon Kass, chairman of the council and a philosopher at the University of Chicago, says he hopes the president will consider both recommendations — while recognizing that one is the majority opinion. "The president asked for a debate of the issues and he's got it," Kass says.

But it is the Congress that must pass any law to ban cloning. And early indications are that the divided recommendation will

entrench the deadlock that has so far prevented the Senate from passing any legislation on the issue.

Gene Tarne, a spokesman for Americans to Ban Cloning, a coalition of anticloning groups, claims that the majority recommendation will help to muster support for a moratorium on human cloning in the Senate. "The four-year period will give time for discussion and encourage scientists to expand work on alternative research," he says.

But Kevin Wilson, director of public policy for the American Society for Cell Biology, which supports cloning for research purposes, says a moratorium would amount to the same thing as an outright ban. He draws comfort from the fact that seven members of a panel named by Bush supported the society's position. "You have a sizeable number of presidentially appointed members thinking that the research should continue," Wilson says.

In a rare display of unity, lobbyists on both sides conceded that the split decision reflected the diversity of opinion among the US public on cloning for research. ■

Poliovirus advance sparks fears of data curbs

Erika Check, Washington

The recent laboratory synthesis of an infectious poliovirus is prompting fears among US biologists that the publication of results deemed potentially useful to terrorists may be restricted.

The virus was synthesized by a team led by virologist Eckard Wimmer at the State University of New York at Stony Brook. The results, published online by *Science* on 11 July (J. Cello, A. V. Paul & E. Wimmer *Science* 10.1126/science.1072266), sparked media reports worldwide suggesting that similar techniques might be used by terrorists to synthesize the Ebola virus or even smallpox.

Geneticists fear that the report — the first demonstration that a published genome can be turned into an infectious virus — will encourage calls for a clampdown on the open publication of gene sequences of viruses. US government officials have been floating the idea of such restrictions since the terrorist attacks last autumn (see *Nature* **414**, 237–238; 2001).

The authors of the paper came under fire from some biologists for failing to offer any defence of the open publication of scientific data — or to discuss the wider context — in their paper. Critics argue that if restrictions on the publication of gene sequences are now implemented, they will cripple research into countering dangerous viruses.

“I’m very worried they’re going to set off a rash of bills that would criminalize scientists,” says Craig Venter, the former president of biotechnology company Celera Genomics in Rockville, Maryland, and now head of the J. Craig Venter Science Foundation, a non-profit group. Venter and other critics say that a harmless virus, rather than a human pathogen, could just as easily have



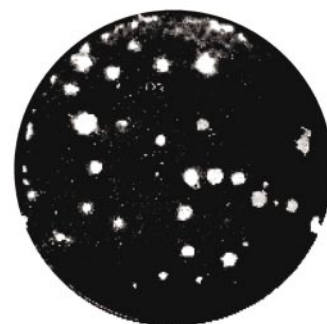
For Eckard Wimmer, his results brought a ‘scary realization’ that viruses can be recreated.

been used to demonstrate the technique.

Wimmer defends his group’s decision to publish, saying that they are only pointing out the fact that terrorists could misuse published data. “We’ve been criticized for playing into the hands of a bioterrorist, but other groups have already written that this could be done,” Wimmer says. He adds that all the information used in the experiment is already publicly available.

Wimmer’s team started with the base-by-base sequence of polio’s genome, which is made of RNA. But making RNA in a test tube is difficult, so the researchers bought DNA corresponding to the RNA from a company that makes synthetic genes on demand. They then used a naturally occurring enzyme to convert this DNA template into RNA. When added to a mixture of human proteins, the RNA copied itself and made new, infectious polioviruses.

Wimmer has previously synthesized infectious polioviruses, but used RNA isolated from wild polioviruses, not from raw



Culture shock: researchers have synthesized this infectious poliovirus in the laboratory.

sequence information. He claims that the latest research proves that the world can never stop vaccinating against polio. “You cannot eliminate a virus from existence,” he says. “Somebody can always resynthesize it, and that’s a scary realization.”

Experts say that it would be difficult — but not impossible — to use Wimmer’s method to synthesize a large and complicated virus, such as smallpox. And because pox viruses carry their own replication machinery they cannot co-opt human enzymes to reproduce themselves. This means that it would be impossible to recreate infectious smallpox using just viral DNA in an extract from human cells, says Geoffrey Smith, a virologist at Imperial College, London.

But Smith and others say that this problem might be solved by adding replication enzymes from a different pox virus to the mixture. And, they add, it would be relatively easy to recreate smaller viruses such as HIV or Ebola using Wimmer’s method. ■

Monkey smallpox trial suspended over painkiller use

Erika Check, Washington

A study of smallpox in monkeys has been suspended by the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, because of concerns about animal treatment.

The study, which began last year, is led by virologist Peter Jahrling of the US Army Medical Research Institute of Infectious Diseases in Fort Detrick, Maryland. Jahrling infected cynomolgus macaques with two strains of smallpox virus in an attempt to develop a monkey model for testing new therapies for the disease. But last week, the CDC suspended the study after Jahrling and veterinarians clashed over when the monkeys should receive painkillers.

The Food and Drug Administration (FDA) ruled in May that animal studies can be used to obtain regulatory approval for drugs (such as smallpox therapies) that cannot ethically be tested on humans. Jahrling says he expects the FDA will require experimental treatments to be given without painkillers, and doesn’t want to press on with a study that might be ineligible for approval.

A CDC spokesperson says that the institution’s animal-care committee had mandated that all the animals in the study should receive painkillers at a certain point, but “word had not gotten to all the vets involved”. The spokesperson says that the agency hopes to resolve the issue at a meeting with institute researchers on 16 July. ■



Peter Jahrling has clashed with vets over when animals in his study should receive painkillers.

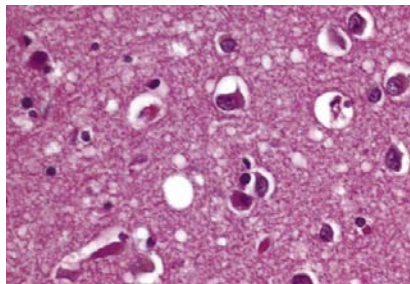
Swiss rise in CJD raises concerns over possible BSE link

London The number of people dying from Creutzfeldt–Jakob disease (CJD) has risen sharply in Switzerland — sparking fears of a possible link with bovine spongiform encephalopathy (BSE).

BSE is thought to be the cause of a distinctive form of the brain-wasting disease known as variant CJD. The Swiss cases, in contrast, are standard ‘sporadic’ CJD. Each year between 1997 and 2000, no more than 11 Swiss people developed CJD. But 19 cases were reported in 2001, and seven were recorded in the first quarter of this year. This is some four times higher than the incidence elsewhere, reports a team led by Adriano Aguzzi of the University Hospital Zurich (M. Glatzel *et al. Lancet* **360**, 139–141; 2002).

The increase could be a mere statistical blip, or it may be due to increased awareness of the disease leading to more diagnoses. More disturbing is the possibility that the cases are linked to the consumption of BSE-infected meat products — which would mean that the BSE agent can cause two distinct forms of CJD.

Possible links between the Swiss CJD cases and BSE will now be explored by



Plaque attack: Swiss patients have spongiform patterns in the brain typical of sporadic CJD.

strain-typing experiments in which the disease is transmitted to mice. These tests will take at least a year to complete. “It’s the best way to establish or exclude any suspected link,” says Moira Bruce of the UK Institute for Animal Health’s Neuropathogenesis Unit in Edinburgh.

Medical journal retracts paper over falsified data

Washington *The New England Journal of Medicine* last week took the unusual step of retracting a paper on heart complications in patients with HIV, after uncovering evidence that the authors had falsified data.

Two of the journal’s editors, Jeffrey Drazen and Gregory Curfman, wrote in its 11 July

issue that they had decided to retract the paper (G. Barbaro, G. DiLorenzo, B. Grisorio & G. Barbarini, *New Engl. J. Med.* 339, 1093–1099; 1998) after finding a suspicious similarity between a microscope-slide image and another image from an earlier paper in *The American Journal of Cardiology*. The authors were asked to retract the paper, but refused to do so.

Drazen and Curfman’s action contrasts strongly with a recent survey by *Nature* of papers implicated in Germany’s biggest case of scientific misconduct, most of which have not been retracted (see *Nature* 418, 113; 2002).

Science lessons send committee to sleep

London British schoolchildren have been saying it for decades — but now it’s official. A parliamentary committee has declared that school science lessons are so “boring” and “overloaded with facts” that they can put people off the subject for life.

The way science is taught to students between the ages of 14 and 16 needs an urgent rethink, says a report from the House of Commons Science and Technology Select Committee. It calls for more flexibility for teachers to discuss contemporary issues, such as genetically modified crops, and longer-term practical projects, such as

investigating local water pollution. The report also calls for greater investment in school lab facilities and technicians, to allow for more interesting practical work.

“School science concentrates too much on churning out exam candidates,” agrees Alistair MacFarlane, chair of the Royal Society’s education committee.

Britain needed over future foot-and-mouth policy

London The emergency vaccination of livestock should be used to control future outbreaks of foot-and-mouth disease in Britain, says a report from the Royal Society.

The report, *Infectious Diseases in Livestock*, also calls for a comprehensive response plan to be drawn up for dealing with any future outbreaks. A similar recommendation was made by the official inquiry into a previous outbreak in 1967–68, but was never implemented.

Six million animals were slaughtered to halt last year’s British epidemic. The government rejected vaccination because officially sanctioned tests could not tell diseased from vaccinated animals.

The Royal Society report notes that tests are now available that can make this distinction. In May, the Paris-based World Organisation for Animal Health agreed to

relax trade restrictions on vaccinated animals that had been checked using such tests.

Sect claims hundreds of cloned embryos

Tokyo Clonaid, a company founded by the Raëlian religious sect, claims to have created hundreds of cloned human embryos that are now stored, ready for implantation into women volunteers. Mainstream researchers are deeply sceptical of Clonaid’s claims, but warn that any attempt to implant a cloned embryo would pose serious risks to the health of both embryo and surrogate mother.

Clonaid representatives made the claim at the 1st International Bio Expo Japan in Tokyo last week, where they were marketing a device for achieving cell fusion, which is central to cloning. The device is produced by Bio Fusion Tech of Daegu, Korea, a company with no other product.

Anglers on line to tackle invading snakeheads

Washington It has been billed in much of the US media as an attack of ‘Frankenfish’ that threatens Maryland’s native freshwater species. In late June, the Maryland Department of Natural Resources received a photograph from an angler of a northern



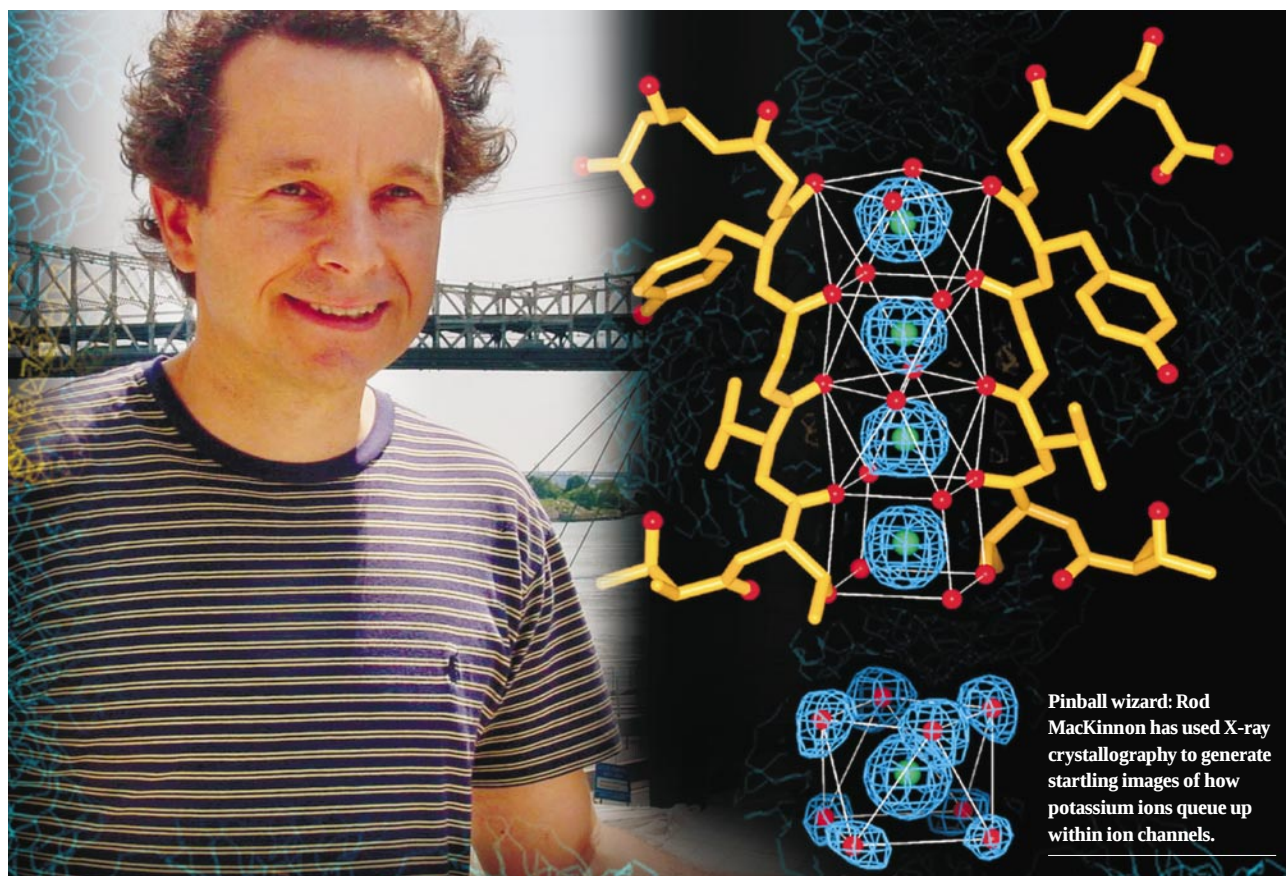
Causing a splash: this juvenile northern snakehead will grow into a voracious predator.

snakehead (*Channa argus*), a voracious predator from China, that was caught in a pond near the town of Crofton. Further reports of catches from the same pond followed. Because the snakehead can survive out of water for several days, and can travel over land for short distances, the incident was soon reported as an ecological disaster.

Investigations revealed that a Hong Kong-born local had released two snakeheads into the pond, which now seem to be breeding. Officials are hopeful that the invaders have not spread and can be eliminated.

► www.dnr.state.md.us/fisheries

Correction In last week’s News Feature on attitudes to misconduct in the physical sciences (see *Nature* **418**, 120–121; 2002), Paul Canfield was misidentified as a materials scientist at the University of Iowa. He is in fact a physicist at Iowa State University.



Pinball wizard: Rod MacKinnon has used X-ray crystallography to generate startling images of how potassium ions queue up within ion channels.

They said it couldn't be done...

Determining the structure of cell-membrane ion channels was thought to be mission impossible. Alison Abbott meets the researcher who proved the doubters wrong, opening new windows on cellular function.

Rod MacKinnon woke up on 1 January 1998 with a sickening feeling that his eureka moment had been a dream. Late into the night, at Cornell University's synchrotron light source in Ithaca, New York, he had been processing data on the structure of a crystallized potassium-ion channel from a cell membrane. Eventually, his colleagues left to join the New Year celebrations, and MacKinnon worked on alone.

Midnight passed, and with each iteration of the data, the image of the channel on his computer screen became clearer. Then, in the channel, shadows of multiple potassium ions began to emerge, lined up like pinballs, exactly as had been predicted some 50 years earlier. "I became so shaky I couldn't hit the keys, and I had no one to tell," MacKinnon recalls. Eventually he went to bed, the excitement of his discovery still buzzing round his head.

Fortunately for MacKinnon, this was no dream. It was, in fact, a stunning highlight among a series of revelations about ion

channels to emerge from his lab. And it was all the more remarkable for the fact that, when MacKinnon embarked on his quest to unveil the channels' structures just a few years before, many structural biologists had regarded him as foolhardy. Didn't he realize that ion channels were almost impossible to crystallize for X-ray structural analysis?

Ion channels are proteins embedded in cell membranes. They act as extremely selective gateways, allowing specific ions to pass in and out of the cell in response to various signals. Most dramatically, they mediate the electrical impulses known as action potentials that are the basis of communication in the nervous system.

I became so shaky I couldn't hit the keys, and I had no one to tell.

MacKinnon didn't start out obsessed with ion channels. After gaining a biochemistry degree, he studied medicine, but soon found that his enquiring mind required a tougher challenge. "I needed to do a real science," he says. So in 1986, MacKinnon returned to Chris Miller's laboratory at Brandeis University in Boston, where he had worked on an undergraduate project, to retrain as an electrophysiologist. This soon lured him towards the mysteries of potassium-channel function.

These puzzles had their roots in the pioneering studies of Alan Hodgkin and Andrew Huxley at the University of Cambridge, UK, who in 1952 showed that a nerve cell's membrane becomes transiently, and very rapidly, permeable to sodium at the start of an action potential¹. Sodium ions rush into the cell, causing the voltage across the membrane to drop. The membrane then becomes permeable to potassium ions, which flow out of the cell, allowing it to return to its resting potential.

Three years later, Hodgkin, with his col-

league Richard Keynes, looked at the flow of radioactive potassium ions across the membrane of a nerve cell. They concluded that ions must be “constrained to move in a single file, and there should, on average, be several ions in the channel at any moment”². MacKinnon got an enormous kick out of finally being able to show that this prediction was correct³. “I love the way Hodgkin thought,” he says.

In the decades before MacKinnon entered the field, Hodgkin and his successors showed that fluxes of ions across cell membranes are mediated by distinct channels, regulated by sensors that respond to particular environmental cues. Some potassium channels, for instance, respond to changes in voltage across the membrane, whereas others react to the presence of neurotransmitter chemicals.

Channelled energy

The problem of exactly how potassium channels open and close became MacKinnon's preoccupation. Having moved to Harvard University in 1989, MacKinnon raced up the academic ladder to become a full professor in 1995. A year later he relocated to New York's Rockefeller University, where he is now a Howard Hughes Medical Institute investigator.

Initially, MacKinnon used electrophysiological techniques to study the properties of potassium channels from bacteria. By mutating the channels at key points and looking at how this changed their properties, he was able to reveal many of their secrets. He showed, for example, that a potassium channel consists of four subunits surrounding a central pore⁴, and was also able to determine which amino acids are responsible for its selectivity — allowing potassium through but blocking other ions^{5,6}.

But these results prompted questions that electrophysiology and mutagenesis could not answer. How are the subunits arranged in three dimensions? What is the precise atomic structure that confers such exacting selectivity? To get the answers, MacKinnon needed to become an X-ray crystallographer. So he started reading books on the subject, and haunting the lab of Harvard crystallographer Steve Harrison. “The postdocs and students thought it was odd, having a professor hanging around, but they got used to it — and they really were my mentors,” says MacKinnon.

Colleagues advised MacKinnon against embarking on what they feared would be an impossible journey. Miller, currently working on sabbatical in MacKinnon's lab, told him bluntly: “You're out of your mind.”

When MacKinnon moved to Rockefeller to begin his new life as a crystallographer, only one of his existing postdocs, Declan Doyle, went with him. “We had a great time,” says Doyle, who now runs his own lab at the University of Oxford, UK. The move coincided with the advent of a highly efficient system for expressing potassium ion channels in bac-



When you are passionate, it is easy to pay attention to all the details.

teria⁷, which offered the crystallographers the large amounts of protein that they needed.

After an initial quiet period, the papers started pouring out. In 1998, MacKinnon and his colleagues published their landmark three-dimensional structure of a potassium ion channel⁸. In the past two years, the researchers have gone on to show how a potassium channel opens by bending its inner helix at a hinge point⁹. They have also revealed the structural elements within certain potassium channels that interact with the sensors that regulate their activity^{8,9} — not to mention two papers last year that revealed the channels' workings in fine detail^{10,11}.

Passionate blend

Why has MacKinnon succeeded where others failed? “Just a dogged determination, and passion — when you are passionate it is easy to pay attention to all the details,” he says. But this success required a lot of care and patience. “Ion channels need to be extracted along with those lipids that support their structure, and crystallized before they dissociate,” MacKinnon says. Fortunately, his experience as an electrophysiologist provided him with the tools — such as toxins that bind to particular sites on a channel — to check at each stage of sample preparation whether the channels are still functional. “You have to keep asking your protein: ‘Are you OK?’” he says.

MacKinnon's highly focused team of ten or so postdocs and students is also central to his *modus operandi*. He endlessly discusses experimental design and the significance of results with lab members — who are very much hand-picked. “When I see promising candidates on paper, I invite them to spend a day in

the lab,” says MacKinnon. “I talk to them for hours, trying to see if they have a scientific mind.” Then he throws them in with his young colleagues to see how they get on. “When you spend so much time together it is important that they have a happy working environment.”

Many subtle ingredients have also contributed to MacKinnon's success. Foremost, perhaps, is the fearlessness that allowed him to leap, inexperienced, at a powerful research tool. “It is better to crash and burn than never to have tried,” he says.

Then there is his love of puzzle-solving. In his leisure time, MacKinnon reads books on the history of mathematics and science in the way that others might read thrillers. “I like to work out how scientists in the past came across the problems they solved, and how exactly they approached their problems.” Suddenly he is seized by embarrassment: “Oh, that may seem a bit nerdy.”

And notwithstanding aberrations such as his lab marathon on New Year's Eve, MacKinnon also tries to keep his life in balance — retreating whenever he can to his summer house in Cape Cod, where he indulges his passions for fly fishing and swimming.

Today, MacKinnon's office, with its fine view of the 59th Street Bridge and the East River, is littered with prizes. Foremost among them is his 1999 Lasker award. Given that around 50% of Lasker winners go on to win a Nobel, colleagues wonder whether MacKinnon will one day emulate his intellectual hero Hodgkin, and receive that call from Stockholm. He has rehearsed his response to such speculation: “What I always tell people is that half of those who get the Lasker don't go on to win the Nobel.”

Ion channels still have some challenges to offer — MacKinnon is keen to crack the secrets of a particularly recalcitrant potassium channel that is regulated by the voltage across the cell membrane. But at the age of 46, he is thinking of another shift in direction. “Once passion is spent it is time to move on, otherwise you start to overlook details,” he says.

What might that new direction be? MacKinnon says he has a few ideas, but for now he is keeping them to himself. Given what he achieved after switching from electrophysiology to structural biology, however, many biologists would dearly love to know what is next on his agenda. ■

Alison Abbott is Nature's senior European correspondent.

- Hodgkin, A. L. & Huxley, A. F. *J. Physiol.* **117**, 500–544 (1952).
- Hodgkin, A. L. & Keynes, R. D. *J. Physiol.* **128**, 61–88 (1955).
- Doyle, D. A. *et al. Science* **280**, 69–77 (1998).
- MacKinnon, R. *Nature* **350**, 232–235 (1991).
- MacKinnon, R. & Miller, C. *Science* **245**, 1382–1385 (1989).
- Heginbotham, L., Lu, Z., Abramson, T. & MacKinnon, R. *J. Biophys.* **66**, 1061–1067 (1994).
- Schrempf, H. *et al. EMBO J.* **14**, 5170–5178 (1995).
- Jiang, Y. *et al. Nature* **417**, 515–522 (2002).
- Zhou, M., Morais-Cabral, J. H., Mann, S. & MacKinnon, R. *Nature* **411**, 657–661 (2001).
- Morais-Cabral, J. H., Zhou, Y. & MacKinnon, R. *Nature* **414**, 37–42 (2001).
- Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. *Nature* **414**, 43–48 (2001).

Can you keep a secret?

Practical products are about to emerge from the weird world of quantum mechanics. Erica Klarreich finds out how quantum cryptography made it from the lab to the marketplace.



The Enigma code machine, used by the German military during the Second World War, was a masterpiece of complexity. Each letter of the alphabet was encoded by a system of wheels that could be set in an almost endless array of configurations, creating a seemingly unbreakable cipher. But Enigma had a chink in its armour.

Messages could not be decoded unless the receiver knew which wheel settings the sender had used. Users exchanged this information, known as a key, at the beginning of the message and encoded it using another, prearranged, key. But security for this second key was not perfect. Enigma users changed it only once a day. After cracking a few messages by focusing on commonly used words, code-breakers could tease out the second key and decipher all of the day's transmissions.

More than half a century later, secret messages are still only as secure as the keys used to encrypt them. Sensitive data are exchanged every day, yet no one has developed a system that ensures that the keys to these messages are absolutely secure.

This may soon change. Companies are close to marketing cryptographic systems that use quantum mechanics to offer absolute security — guaranteed by the laws of physics. Devices that can securely transmit keys through fibre-optic cables may soon be available. And it might eventually be possible to transmit quantum keys using satellites, allowing users across the world to form secure connections.

The keys used to encrypt most messages, such as those used to exchange credit-card information over the Internet, are themselves encrypted before being sent. The schemes used to disguise keys are thought to be secure, because cracking them would take too long for even the fastest computers. The widely used RSA algorithm is one example. Anyone wanting to receive a message pub-



Code created by the Enigma machine was eventually cracked, but unbreakable codes could soon be carried down optical fibres (inset) by single photons.

lishes two numbers, one of which is the product of two very large prime numbers. Senders convert their message into a series of digits, and perform a simple mathematical calculation on the series using the publicly available numbers. Messages are deciphered by reversing the calculation.

Prime movers

This is easy to do if you know the values of the prime numbers, which are not published. An eavesdropper can, in principle, work them out, but for big enough numbers this would take millions of years with the computing power available today.

The future performance of such systems depends on estimates about the speed of future computers, and such guesses have proved wrong in the past. In 1977, for example, *Scientific American* challenged computer scientists to decode a message encrypted using a 129-digit number¹. Ron Rivest, a computer scientist at the Massachusetts Institute of Technology and one of RSA's creators, estimated that it would take

4×10^{16} years to factor such a number. But in 1994, a team of computer scientists and amateur volunteers managed to decipher the message by applying 1,600 computers to the problem over a period of eight months².

In the same year, computer scientist Peter Shor of AT&T Labs in Florham Park, New Jersey, described a new kind of threat. Shor was interested in quantum computers — hypothetical machines that should be able to carry out a large number of calculations simultaneously. He showed that if such a computer is ever built, it will be able to factor large numbers rapidly, and could quickly crack all the commonly used public key systems³.

A quantum computer or new factoring technique might not come along for decades. But some secrets encrypted today, such as the design of nuclear weapons, will still be important then. "We have to assume that any information encrypted today is probably being recorded by eavesdroppers in the hope that it will be of value 10 or 20 years into the future," says Richard Hughes, a physicist at Los Alamos National Laboratory in New

Mexico who works on quantum cryptography. "If they have quantum computers, they'll be able to look at information encrypted today and learn useful things from it."

But while quantum computers remain no more than an interesting possibility, another quantum technology could soon be ensuring total security. In quantum mechanics the act of measurement changes the properties of the very thing being measured. This is a boon to cryptologists, because it means that eavesdroppers cannot listen to certain types of information without leaving an unmistakable disturbance.

Polarized opinion

The details of a quantum cryptography system were first described in 1984 by theoretical physicists Gilles Brassard of the University of Montreal in Canada and Charles Bennett of IBM's Thomas J. Watson Research Center in Yorktown Heights, New York⁴. In their scheme, Alice sends Bob a series of ones and zeros, which are used to generate the key. Each 'bit' is represented by a photon of light with one of four possible polarizations: horizontal, vertical or one of the two diagonals. Alice and Bob agree that a horizontal polarization corresponds to a zero and a vertical polarization corresponds to a one, and make a similar decision for the two diagonal polarizations.

Quantum mechanics says that Bob can either look to see whether the photon is horizontally or vertically polarized, or which of the diagonal polarizations it has, but he cannot do both. When a photon arrives at Bob's end, he randomly chooses which of the two types of orientation to test for. If Alice has sent a vertically polarized photon and Bob makes a horizontal-vertical measurement, he will discover the polarization correctly and read off a one.

But if Bob makes a diagonal measurement, the vertical polarization of the photon lies exactly midway between the orientations he is looking for. The photon will instantly switch to one of the diagonal polarizations, each with the same probability, and Bob has an equal chance of recording a one or a zero. Bob will also get a random result if he makes a horizontal-vertical measurement on a diagonally polarized photon.

At the end of the transmission, Bob knows he has correctly measured the polarizations of about half of the photons, but doesn't know which ones. So Bob contacts Alice on a channel that doesn't have to be secure, and tells her which type of measurement he made for each photon, but not the outcome of those measurements. Alice tells him which measurements were correct. They discard the incorrectly measured photons, and keep the rest for their key.

To make sure that a third party, Eve, hasn't listened in to their original exchange, Alice and Bob next sacrifice a small amount of their



Light work: keys encoded using polarized photons have been sent between Alice and Bob (left) through 67 km of fibre-optic cable under Lake Geneva.



key and check it over the public channel for errors. If Eve has been assessing the polarization of the photons somewhere between Alice and Bob, she will have changed the polarization of about half of them. This makes about one in every four entries in Bob's key different from those that Alice sent — a clear indication that Eve has been snooping. And there is no way consistent with the laws of physics for Eve to cover her tracks.

In reality, noise in the channel through which the photons pass introduces a small number of errors into the transmission. So a clever eavesdropper could gain some information about the key by measuring such a small number of photons that Alice and Bob cannot distinguish the errors this introduces from those caused by noise. Alice's single-photon generator could also cause problems by occasionally sending out two photons instead of one. Eve can divert and measure one of the photons, while allowing the other to proceed to Bob. But in each case, Bob and Alice can generate a new key by applying an algorithm to their existing key. Eve, who is missing the bulk of the original key, cannot

hope to predict the outcome of this algorithm.

In 1989, a team led by Bennett and Brassard built a working device, and sent photons through the air to a receiver about 30 centimetres away⁵. By the mid-1990s, other groups were sending encrypted keys through tens of kilometres of optical fibre. And over the past few years, the first steps towards commercializing such systems have been taken. "Quantum cryptography is very much a reality," says Brassard.

Gone in a flash

Last October, a group of physicists at the University of Geneva in Switzerland launched a company called id Quantique, which will supply a system integrating the cryptography hardware — photon sources and detectors, and fibre-optic connections — needed to exchange keys. In March this year, they used the system to send single photons through 67-km telecommunication cables running under Lake Geneva⁶. "The system is very stable, and has the potential to be very fast," says Nicolas Gisin, a member of the team.

MagiQ Technologies, a New York firm that specializes in quantum technologies, is trying to build a system in which the discussion about which photons have been received correctly is streamlined and integrated with the photon generators, detectors and fibre-optics. The firm hopes to market a full quantum-cryptography system by early next year.

MagiQ and id Quantique's systems are designed to connect users who are linked by a single dedicated fibre, but other groups are working on systems that can support a network of users. Last September, BBN Technologies, an information-technology company based in Cambridge, Massachusetts, began a five-year collaboration with teams at Boston and Harvard universities to build



Gilles Brassard (left) and Charles Bennett laid the foundations of quantum cryptography.



In the air: Richard Hughes has sent a photon-encrypted code from a laser source (circled, inset) to a receiver.

a quantum network connecting the three institutions. Photons will be routed round the network using mirrors. "The mirrors send the photon along without measuring it, so they don't create the kind of disturbance an eavesdropper would," explains Chip Elliott, an engineer at BBN.

Working devices may soon be on the market, but that does not mean that the engineers involved can rest on their laurels. Reliable single-photon generators, for example, are not yet commercially available. Today's systems, such as those developed by id Quantique, instead use lasers that generate pulses so weak that they almost never contain more than one photon. But at such low intensities, nine out of ten attempts to fire a photon fail.

Current photon detectors also present some problems. To spot a single photon, the detectors must be so sensitive that they will sometimes register photons that are not there. Even then, they will typically miss 90% of the transmitted photons. What's more, many photons are absorbed by the optical fibre and never make it to the receiver. "We send five million bits per second, but by the time we get done with all the detectors and the specialized protocols that shorten the key during the public discussion, we get somewhere between 100 and 1,000 bits per second," says Elliott.

But this is enough for cryptographic uses. The Advanced Encryption Standard, the encryption algorithm used by the US government, uses a key with a maximum of 256

bits. A key distribution that sends 500 bits per second would allow users to change the key roughly twice per second, more than ample for most purposes.

The distance that the key can be transmitted is a more important technical limitation. Most experts agree that the Geneva group's 67-km transmission is close to the maximum that can be achieved with current technology. Beyond about 80 km of cable, too few photons make it from Alice to Bob. Both id Quantique and MagiQ are reluctant to discuss who is interested in their products, but this limitation means that the first users are likely to be organizations that want to transfer highly secret material within a single city, such as government offices, banks and businesses.

Long-range forecast

The range could be extended by devices that strengthen the signal as it passes by, like those used to send telephone conversations over long distances. But unlike telephone repeaters, quantum versions would have to bolster the signal without measuring the photons. "A repeater that doesn't measure was thought to be impossible in the early 1980s, but since then scientists have shown that it is feasible in principle," Brassard says. "But we're nowhere near the technology to build one."

Satellites could provide an alternative means of achieving long-distance transmission. Hughes' team at Los Alamos is developing a key-distribution system that sends single photons through open air. So that the photons can be distinguished from all the

others bombarding the detector, the team uses various techniques to filter the incoming light. The detector only accepts photons within a narrow range of wavelengths — about 0.1 nanometres — and ignores photons that arrive from angles outside a window of about a hundredth of a degree. A bright pulse of light is also sent 100 nanoseconds ahead of each photon, cueing the detector to expect the next signal.

"When we threw in these three filters, we could get the amount of light down to the level where we could detect the photons we wanted, even if the Sun was shining directly on the receiver," says Hughes. In a paper published this month⁷, Hughes and his colleagues describe how they sent keys over a distance of 10 km with rates similar to those achieved using optical fibres.

Ten kilometres is still a long way short of the hundreds of kilometres between the Earth's surface and satellites. But because air turbulence, the factor that most disrupts the photons, occurs predominately in the lower two kilometres of the atmosphere, Hughes believes his system should be able to send signals to satellites. "I don't see any show-stoppers at all to doing this from ground to satellite," he says. The team is now trying to make the receiver light and sturdy enough to fit in a satellite and survive a rocket launch.

Combined with optical fibres, satellites could eventually form part of a long-distance transmission system. In the shorter term, the technology might help to protect the security of satellite television broadcasts. In one embarrassing breach, a hacker known as Captain Midnight interrupted a 1986 broadcast by US company Home Box Office and sent over half of the company's customers a five-minute broadcast of a message complaining about the firm's new subscription charges.

Quantum cryptography may soon be helping to prevent similar lapses, and to protect sensitive transmissions. Within the next few months, such systems could start encrypting some of the most valuable secrets of government and industry. Cryptography is about to lose its Achilles' heel. ■

Erica Klarreich is journalist in residence at the Mathematical Sciences Research Institute in California.

1. Gardner, M. *Sci. Am.* **237**, 120–124 (1977).
2. Atkins, D., Graff, M., Lenstra, A. K. & Leyland, P. C. in *Advances in Cryptology — ASIACRYPT '94* (eds Pieprzyk, J. & Safavi-Naini, R.) 263–277 (Springer, Heidelberg, 1995).
3. Shor, P. W. in *Proc. 35th Annu. Symp. Foundations Comp. Sci.* (ed. Goldwasser, S.) 124–134 (IEEE Computer Society Press, Los Alamitos, California, 1994).
4. Bennett, C. H. & Brassard, G. in *Proc. IEEE Int. Conference on Computers, Systems & Signal Processing* 175–179 (IEEE Press, Los Alamitos, California, 1984).
5. Bennett, C. H., Brassard, G., Salvail, L. & Smolin, J. *J. Cryptol.* **5**, 3–28 (1992).
6. Stucki, D., Gisin, N., Guinnard, O., Ribordy, G. & Zbinden, H. *New J. Phys.* **4**, 41 (2002).
7. Hughes, R. J., Nordholt, J. E., Derkacs, D. & Peterson, C. G. *New J. Phys.* **4**, 43 (2002).

♦ www.idquantique.com

♦ www.magiqttech.com

♦ www.bbn.com

Names: a historical or political perspective?

Sir — I have two questions in response to Rognes's Correspondence (*Nature* **417**, 379; 2002). First, has most of his research work been done before or after June 1967? The "occupied territories" have been within Israel's borders since that date. If the major piece of research was performed after 1967, surely the territories should be referred to by their current names? Would anyone in the scientific community nowadays call India a part of the 'British Dominion', or place Ukraine in the 'USSR'? Would anyone use the name 'Rhodesia' when referring to Zimbabwe or 'Ceylon' instead of Sri Lanka?

As with any scientific nomenclature, a scientist is required by a journal to use the current names of territories, regardless of his or her politics. It may be useful to add comments regarding historical names if they are relevant to the research described.

Second is the issue of showing some respect to the publisher's country. If you know that certain names are a sensitive issue in certain countries, and you do not wish to take sides in a political debate local to the area, why not use the terms that are used by the country concerned and by most of the world? I do not think *Nature's* publishers in the United Kingdom would be happy to review manuscripts referring to the Falkland Islands as the 'Malvinas'.

Sharona Even-Ram

National Institute of Dental and Craniofacial Research, National Institutes of Health, 30 Convent Drive, Bethesda, Maryland 20892, USA

Nature allows contributors to use their own choice of name (if in standard use) for the region from which they are writing. This holds for the Falkland Islands/Malvinas as for other regions of the world, such as Taiwan — Editor, Correspondence

Only vital need justifies primate experiments

Sir — In your recent Opinion article on when experiments with non-human primates can be justified (*Nature* **417**, 684–687, 2002), a relevant perspective can be found in US constitutional law. A non-human primate can be regarded as belonging to the 'quasi-person' class originally used in 1865 to define African-Americans, who were not accorded full constitutional personhood in law, but were entitled to some natural rights under the US constitution, including that of self-interest (see R. M. Lebovitz, *St Thomas Law Review*, **14**, 561–600; 2002). But does this mean that primates are entitled

constitutionally to be free from the physical abuse produced by experimentation?

No constitutional right is absolute, but there must be an important and persuasive reason for abrogating it. In the United States, a person's so-called 'liberty interest' is a fundamental constitutional right, yet the government can infringe upon it when the circumstances are important enough. A famous example is an early twentieth century case in which an adult man refused to be vaccinated against smallpox, despite a Massachusetts state law requiring it. The case made its way to the US Supreme Court, where the justices wrote that constitutional rights are not unencumbered, but may be infringed when necessary for the good of the community, including for reasons of public safety to protect the population from a disease epidemic (Jacobson vs Massachusetts, 197 US 11; 1905).

Using the same rationale, it can be argued that, although most animal experimentation is not significant enough to warrant violating a primate's right of self-interest, some experimentation may meet this high standard. For instance, research on a particular vaccine may be important enough as a societal concern to excuse intrusion on the animal's right of self-interest. We must carefully scrutinize the goals of experimentation to determine

when it justifies extinguishing a US animal's constitutional right of self-interest.

Richard M. Lebovitz

Interdisciplinary Bioscience PhD Program, George Mason University, 4400 University Drive, Fairfax, Virginia 22030-4444, USA

Little funding to develop non-animal testing

Sir — Your News Feature "The great primate debate" (*Nature* **417**, 684–687; 2002) did not discuss the potential of *in vitro* and clinical studies to replace many primate experiments in drug development, neuroscience and AIDS research. Non-animal methods can — and should — be supported by all sides of the debate. While verbal support is often forthcoming, resources to develop such alternative methods are not. Despite its responsibility under European legislation, the British government's budget for alternatives is sub-divided into different areas, with the intention that developing replacements for animal experiments should receive about £56,000 (US\$85,500) a year — barely enough to fund one research project.

Gill Langley

Dr Hadwen Trust for Humane Research, 84a Tilehouse Street, Hitchin, Herts SG5 2DY, UK

Are results of primate research worth the suffering it causes?

Researchers must be honest about methods and goals.

Sir — Your Opinion article ("Distasteful but necessary"; *Nature* **417**, 673; 2002) does not explore in depth the morality of experimenting on non-human primates. You claim that the issue is "simple", that "potential benefits must be weighed against the suffering caused". But nobody would apply a cost-benefit test to experimenting on people against their will. What is the difference? Your argument can only be that non-human primates are one sort of primate and we are another, so it is acceptable for us to cause them pain.

That approach cannot withstand any intellectual scrutiny. Primates are just as capable of suffering as we are. Suffering, and its avoidance, lies at the heart of all moral philosophy. If it is wrong to cause suffering to people deliberately — whatever the benefits to others — it must also be wrong to do so to other primates.

You also say that it is time to inform public opinion about primate research. I agree. But that has to mean full

information, not simply such information as researchers find it convenient to disclose. It should not take undercover investigations, such as that recently carried out by the British Union for the Abolition of Vivisection (BUAV) into primate research at Cambridge University, to reveal the extent of suffering involved (caused both by the experiments and the housing conditions) and to raise serious questions about the usefulness of the research.

BUAV is calling for an inquiry into primate research in the United Kingdom independent of the Home Office, which has once again shown, in the Cambridge case, that it is not up to its regulatory task. Will primate researchers support BUAV in this call? If not, it will only strengthen the growing perception that researchers and government alike would prefer to keep the public in the dark.

Michelle Thew

British Union for the Abolition of Vivisection, 16a Crane Grove, London N7 8NN, UK

Viewing life as cooperation

Can symbiosis and genome acquisition account for all speciation?

Acquiring Genomes: A Theory of the Origin of Species

by Lynn Margulis & Dorion Sagan

Basic Books: 2002. 256 pp. \$28, £16.95

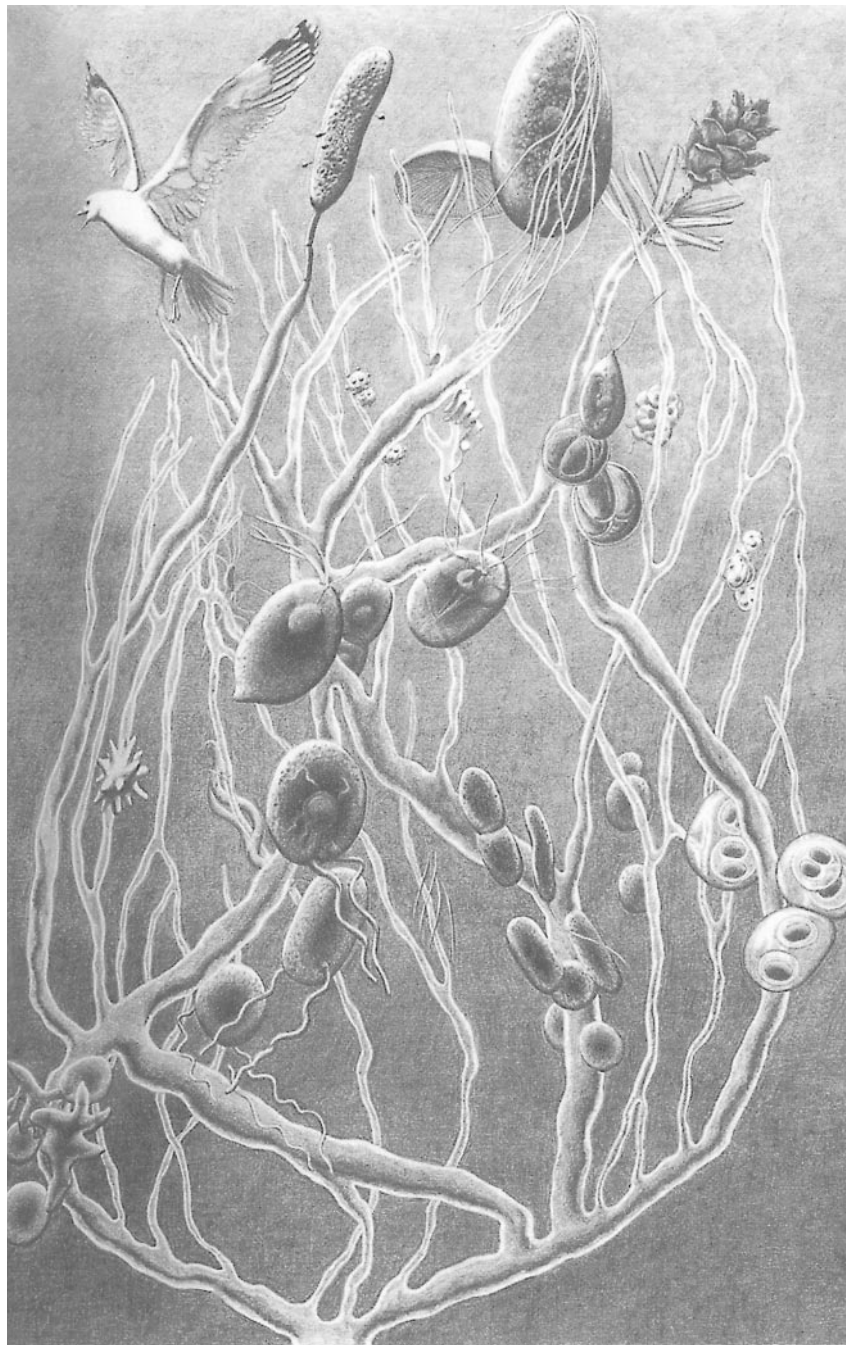
Axel Meyer

Linus Pauling is reported to have said that one needs to have many ideas to have a few good ones. Lynn Margulis has had many ideas, some of the most heretical of which were initially thought to be plainly wrong. Some of them turned out to be brilliant insights that changed current thinking on the early history of life and the origin of the eukaryotic cell. After long battles, Margulis convinced biologists that eukaryotes originated through the fusion long ago of an archaeobacterium with some eubacteria. It is also now well accepted that mitochondria and the chloroplasts of plants were once independently living prokaryotes. We also know that corals, lichens and cows, as well as many other familiar life-forms, are the products of integrating symbionts.

Many less-familiar species in the microbiological world took symbiosis events to the next level of integration: true genomic unions. These cannot be broken up again and have to follow joint evolutionary trajectories. Furthermore, it is well established that, early in the history of life, combination of genomes or portions of genomes took place through lateral gene transfers. As a result, establishing the 'tree of life' through DNA sequences is all the more challenging.

In *Acquiring Genomes*, Lynn Margulis and Dorion Sagan apply their ideas of a symbiotic origin of life to a theory that all speciation is not due to random events and accepted neodarwinian processes, such as mutations and natural selection acting through competition and sexual selection on variants among members of one species. Rather, they argue that all speciation events are caused by symbioses, cooperation and the reticulation of genomes, questioning some of Charles Darwin's central ideas. In their view, Darwin was wrong to emphasize competition and selection as the sole forces shaping the origin of species; they postulate instead that cooperation and symbiosis drive evolution.

I think it is a remarkable tribute to Darwin that almost 150 years after his *Origin of Species* was published, it is still seen as the benchmark against which many authors are compared. It attests to the profundity of Darwin's insights that new ideas are still pitted against his words from two centuries ago. What other scientific discipline can look back at the work of a single individual whose



Life changing: widespread genome acquisition would radically affect our view of the tree of life.

main work was published so long ago and yet is still vigorously debated today? Why should one even expect that Darwin could have been correct all the time? Of course, genes, genomes and any form of molecular biological knowledge were unknown to Darwin; how then could his ideas even begin to explain all that we know about evolution today?

In this age of genomics we are forced to re-evaluate some of his ideas in light of the new data on the genetic variation and conservation among individuals, species, phyla and kingdoms. However, I believe that Margulis and Sagan are wrong to postulate that speciation is driven by symbiogenesis (defined as long-term stable symbiosis that

CHRISTIE LYONS

leads to evolutionary change), rather than by the time-honoured processes that have been amply documented since Darwin's day. Surely, symbiosis was important in evolution, and possibly in some instances of speciation, but these events only account for some of the exceptions to the rule.

Margulis and Sagan go so far as to define species as sets of individuals that are composed of the same set of integrated genomes. They argue that species are unique products of symbiogenesis, and postulate that "no visible organism or group of organisms is descended from a single common ancestor". Consequently, because no bacteria (either eubacteria or archaeobacteria) evolved from the genomic integration of formerly independent bacteria, they must lack species altogether. The biological-species concept and several alternatives have difficulty in defining species among bacteria or unisexual species, but most researchers would not doubt that bacteria have independent evolutionary lineages that are justly considered to be species. Margulis and Sagan also argue that the primary level of selection is not the gene, the individual or the species, but rather the cell.

In his foreword to the book, Ernst Mayr, while supporting the general message that genomes have been acquired, warns of trying to stretch a good idea too far. He cautions that the authors "sometimes arrive at interpretations others of us find arguable" and says: "let the readers ignore those [ideas] that are clearly in conflict with the findings of modern biology." Mayr is pleased that Margulis and Sagan counteract reductionist tendencies in modern biology, which focus exclusively on the level of the gene, by pointing out that some organelles (such as mitochondria and chloroplasts) were once a different species of bacteria. As a result, they argue, the organism as a whole needs to be considered in attempts to understand how it came into being, and how it is functioning now.

It is my conviction that the history of a species and each of its members is laid down in its genome, allowing us, at least in theory, to reconstruct its entire evolutionary lineage. Despite such complicating phenomena as horizontal gene transfer, transposable elements and viral integration of genes, we can generally read from a species' genome how, when and from what a species evolved. Admittedly, things are more messy nearer the bottom of the tree of life, where evolutionary lineages crossed and fused probably on a larger scale than today. The events emphasized by Margulis and Sagan are probably occasional yet important cases of massive reticulate evolution, but today the expected vertical rather than horizontal mode of evolution prevails.

The authors try with admirable single-mindedness to convince the reader that their

idea of symbiogenesis applies to most biological phenomena. Some of their fervour is surely overzealous; despite this, or perhaps because of it, *Acquiring Genomes* is one of the most stimulating and provocative books that I have read for a long while. ■

Axel Meyer is in the Department of Biology, University of Konstanz, Konstanz D-78457, Germany.

Unravelling a heavenly mystery

The Biggest Bangs: The Mystery of Gamma-Ray Bursts, The Most Violent Explosions in the Universe
by Jonathan I. Katz
Oxford University Press: 2002. 218 pp.
£14.99, \$28

Flash! The Hunt for the Biggest Explosions in the Universe
by Govert Schilling, translated from Dutch by Naomi Greenberg-Slovin
Cambridge University Press: 2002. 320 pp.
£18.95, \$28
Luigi Piro

The revolution that has transformed γ -ray bursts (GRBs) from an obscure riddle into one of the hottest topics in astrophysics is only a few years old. Discovered in the late 1960s, GRBs are intense flashes of γ -rays that appear suddenly in the sky, about once a day, from random and unpredictable directions. They typically last for a few seconds and then vanish below the detection threshold of γ -ray instruments. The limited ability of these detectors to identify exactly where in the sky the bursts were occurring meant that it was not until 1997 that GRBs were linked to events occurring at any other wavelengths and their distance scale was calculated.

Measurements made by the Burst and Transient Source Experiment (BATSE) onboard the Compton Gamma-Ray Observatory, launched in 1991, showed that GRBs were scattered across the sky, with no significant clustering in any direction. This suggested only two plausible explanations: that GRBs were produced by a previously unknown population of neutron stars in the halo of our Galaxy (a galactic origin) or that they were a result of extremely powerful stellar explosions in distant galaxies (a cosmological origin).

The launch of the Italian–Dutch Beppo-SAX satellite in 1996 opened a new window on space. Its instruments and ground operations were able to deliver fast, precise positions of GRBs and, by rapidly reorienting the satellite, it was possible to search for any faint post-burst X-ray afterglows using a set of highly sensitive X-ray telescopes. On



In with a bang: the explosion of a massive star led to the formation of the Crab Nebula.

28 February 1997, the first X-ray afterglow was detected, a faint, fading source marking the location of the GRB that had occurred eight hours previously. This and later detections of afterglows at optical and radio wavelengths enabled researchers to conclude that GRBs originate in galaxies that are billions of light years away from Earth, at distances similar to those of the most distant quasars. The energy output inferred from the intensity observed at Earth is 10^{51} – 10^{54} erg, making GRBs the brightest sources in the Universe. In the past few years, the study of afterglows has provided evidence that GRBs are associated with star-forming regions and possibly with supernovae. We now know that GRBs and their afterglow emission are produced by a relativistic expanding fireball — a cauldron of γ -rays, electron–positron pairs and protons in which matter is blown away at almost the speed of light by the enormous energy released in the initial explosion. And investigations are currently under way to discover the central source of the explosion and its progenitors.

The Biggest Bangs by Jonathan Katz and *Flash!* by Govert Schilling are popular-science books that tell the story of the investigation into GRBs. Both introduce the reader to the theory and observations of these objects in layman's language, giving an inside view of how scientists set about unravelling the mystery of GRBs. And they show clearly how a scientific revolution is never the result of a simple linear progress of theory and observations — it is always

accompanied by false starts, controversies, intuition, science politics and the stubbornness and enthusiasm innate in we humans.

But there are significant differences between the two books. *The Biggest Bangs* details the early measurements, achievements, false trails and problems that scientists faced before afterglows were discovered. Katz is a theoretician who has worked in the GRB field, and he turns a physicist's eye on the main astrophysical problems using non-mathematical yet rigorous prose. Often, he widens the problems encountered in the GRB field to touch on other major riddles in modern astronomy. His explanations are enriched with informative historical links. Take, for example, the case of the compactness paradox: GRBs appear to be so small and so luminous that the γ -rays we observe should not be able to escape the source. In fact, the solution to this conundrum lies in the properties of the relativistic expanding fireball, and the foundations for this discovery can be traced back to 1939 and the special properties identified in the fluctuating stars called novae, and to the 1970s with quasars.

The section of the book describing recent breakthroughs contains some historical inaccuracies and is disappointingly short, taking up about a fifth of the book, with only three figures. This is too brief to give a proper account of the observational and theoretical developments of the past years. For example, although the discovery of the first radio afterglow and its rapid fluctuations are described, there is no discussion of the most important implication of this observation — that it allows the speed of expansion and the size of the radio source to be measured, and provides the most elegant and direct proof of relativistic expansion.

These aspects are covered in *Flash!*, which deals primarily with recent discoveries — the X-ray, optical and radio afterglows, the identification of GRBs as being of cosmological origin and their connection with supernovae and star-forming regions in distant galaxies. Future missions devoted to GRB observations are also covered. Schilling, a professional science writer, skillfully leads the reader through the work of scientists in the field, giving a behind-the-scenes view based on interviews with most of the protagonists, and a clear description of the basic concepts behind the theory, observations and instruments. The personal side is somewhat overemphasized, but as a whole the book is an excellent account of research in the field.

Both of these books provide a highly readable account of the story of GRBs, and each in its way is an enlightening and stimulating read for the layman. ■

Luigi Piro is in the Istituto Astrofisica Spaziale e Fisica Cosmica, CNR, Via Fosso Cavaliere, 00133 Rome, Italy.

Darwin's personal voyage

Fossils, Finches and Fuegians: Charles Darwin's Adventures and Discoveries on the Beagle, 1832–1836

by Richard Keynes

HarperCollins: 2002. 428 pp. £25

Keith Thomson

"After having twice been driven back by heavy south-west gales, His Majesty's Ship *Beagle*, a ten-gun brig, under the command of Captain FitzRoy, R.N. sailed from Devonport on the 27th of December, 1831." This first sentence from Darwin's *The Voyage of the Beagle* still works its magic. In fact, *HMS Beagle*, a 90-foot brig re-rigged as a barque, made three major surveying voyages: to the coasts of southern South America (1826–30); a return to South America and then the first complete circle of meridional measurements around the globe (1831–36); and the first full survey of the coasts of Australia (1837–43).

It was for the second voyage, now known as *the voyage*, that Captain Robert FitzRoy took the young Darwin along as a naturalist. FitzRoy originally hoped to share with Darwin authorship of the official *Narrative of the Surveying Voyages of His Majesty's Ships Adventure and Beagle, 1826–1836*. But Darwin preferred to produce his own separate account, making a third volume (a fourth consists of tables and memos). Darwin's *Journal of Researches* soon outsold the rest and continues in print as *The Voyage of the Beagle*.

Commentaries on the voyage abound, among the best being *Darwin and the Beagle* by Alan Moorehead (Hamish Hamilton, 1969) and a previous book by Keynes, *Beagle Record* (Cambridge University Press, 1979).

Both bring a wealth of supplementary information and illustrations to the original (Darwin seems never to have got the hang of using pictures). Now Keynes has returned to the subject to re-tell Darwin's tale in even more detail (although the new book is not as well illustrated and has a modest index).

As is perhaps forgivable in a book by a great-grandson of Darwin, this is a conventional hagiography. There is no hint of anything that might reflect badly on Darwin (such as the bitter row with FitzRoy over Darwin's failure properly to acknowledge in his book the help of his shipmates) or that is even mildly controversial. Darwin's later illness and the argument over why he was on board in the first place are glossed over. Two madly giddy letters to Darwin from FitzRoy are presented as if perfectly matter-of-fact, and FitzRoy's mental breakdown in Chile is downplayed.

Setting aside the enigmatic genius of FitzRoy and the tragedy of the Fuegians whom FitzRoy kidnapped in 1830 and returned to Tierra del Fuego in 1833, Darwin's book continues to fascinate because it is a wonderful read and a source of clues to the 'real' Darwin and the sources of his ideas. After quitting medical school in Edinburgh in 1828, he was accused by his father of caring only for "shooting, dogs, and rat catching and [you] will be a disgrace to yourself and all your family". And in 1831 Darwin graduated from the University of Cambridge with a pass degree, having prepared for the Church. Yet by 1837 he was an acknowledged scientist and already had the germ of his theory. How did he use a shipboard library and five years of exploring, collecting and observing to turn himself into an intellectual?

The answer is largely that he was no playboy but a serious young man in search of a defining role. Although Darwin noted in his autobiography that the voyage was "the most



On his travels Darwin encountered the Fuegians, depicted here by the *Beagle*'s artist, Conrad Martens.

important event of my life and has determined my whole career", he did not reveal, and we still do not know, precisely why and how. The main unresolved issue concerns the transmutation of species, a subject with which Darwin had been familiar since his studies in Edinburgh and from his grandfather's *Zoonomia*, plus his reading during the voyage of Charles Lyell's *Principles of Geology*. Nine months after his return, Darwin wrote that he had started thinking seriously about the subject in March 1836, when the ship left Australia to start the long (and for Darwin, contemplative) trek home. The immediate triggers were fossils and species on the Galapagos Archipelago, which he visited in September and October 1835.

Keeping things strictly chronological, Keynes writes brilliantly about what happened when and where, and how Darwin wrote it up in diaries, notebooks and manuscripts. But there is little mention of the enormous literature on the voyage, and in the end Keynes has not shown us anything new about the development of Darwin's philosophical thinking on questions of creation or transmutation over the period from 1831 to 1836. The result is a work of great erudition and a valuable addition to the literature, but the principal questions remain, as I suspect Darwin always intended, unanswered. ■

Keith Thomson is at the Oxford University Museum, Parks Road, Oxford OX1 3PW, UK, and is the author of *HMS Beagle* (W. W. Norton, 1995).

Science in culture

Leonardo's layer?

A scientific examination of Leonardo da Vinci's *Adoration of the Magi* turns up some surprises.

Martin Kemp

Science has revolutionized what we can see in paintings. Non-invasive techniques bear witness to otherwise invisible changes of mind (using X-rays), disclose hidden underdrawings (with infrared rays), and detect pigments that were added to the painting after its completion (through infrared and ultraviolet illumination). X-ray fluorescence provides analyses of the inorganic pigments, and photometry measures the spectrum of the reflected light. Gas chromatography identifies organic materials such as binding media and varnishes from tiny pigment samples, and microscopic cross-sections reveal the stratigraphy of the paint and varnish layers.

The new data clearly provide a great resource for art historians. However, as with any body of evidence, we need to learn how to see what is significant in the visual output and how to interpret it in the context of other types of established knowledge in the field. Science does not provide absolute answers for the art historian, but it allows us to ask ever more complex questions about the physical composition of pictures, and propose solutions with increasing confidence.

The context for these remarks is the exciting evidence beginning to emerge from the technical examination of Leonardo da Vinci's famous *Adoration of the Magi* (below), which hangs in the

Uffizi Gallery in Florence, Italy. The altarpiece was subject to a complex agreement with the monks of San Donato a Scopeto in 1481, which decreed that the painter should receive a portion of the deceased patron's country property and deposit a sum in the Florentine dowry bank in favour of the patron's granddaughter. Unhappily, the painting on the large square panel had not progressed beyond an underpainting when Leonardo left for Milan, probably in 1482, having also defaulted on his obligations to the young lady.

Infrared reflectography, which can disclose carbon-based underdrawing on the white gesso priming, had already revealed more of the subtle beauties of Leonardo's touch than is visible on the discoloured surface of the present picture. The obvious question arose: should the painting be cleaned? When the possibility was made public, inevitable controversy ensued. Maurizio Seracini in Florence, who had previously provided the infrared data, was then asked to undertake a full-scale technical examination.

On my recent visit to Florence, Seracini was kind enough to provide a briefing on what he is discovering. The most surprising implications arise from his microscopic examination of cross-sections of the paint and varnish layers. As yet these are too scattered to allow definitive judgements, but his observations indicate that we will have to revise our thinking about the painting.

The cross-sections suggest that Leonardo unprecedentedly laid a semitranslucent layer of white lead over his delicate underdrawing, which had been undertaken with a fine brush for its linear design and a broader brush with diluted pigment for the shading. On top of this layer he added the central trees, but he abandoned the picture before completing it with any final strata of fully coloured pigments.

Seracini also questions the authenticity of the heavier brownish layer that begins to establish the background of shadow from which some of the figures emerge like ethereal spirits. It seems likely from the cross-sections that this paint layer is not part of Leonardo's original structure, not least because it has seeped into cracks in the lower strata — implying that it was added much later.

But nothing is ever simple where Leonardo is concerned. So little is known about the history of the panel for almost 300 years after it was painted that we have no explanatory model for when, where, by whom or for what purpose the brownish pigment might have been added, if it were not done by Leonardo himself. One question that I believe is nearer to an answer is whether the painting should be cleaned. It looks as if the layers that are definitely by Leonardo are integrated in such a complex way with those that might not be by him that we would be well advised at this point to leave well alone.

Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.



Weaving life's pattern

Melvin Konner

Psychologists like to stress that what happens in early life — what zoologists call the juvenile phase — is not just growth, but development. The implication is twofold. First, 'growth' suggests mere augmentation, either through increasing cell size (hypertrophy) or successive mitotic divisions (hyperplasia). But a system such as the brain could not emerge so simply. Second, 'growth' implies an autonomous process, governed from within, and (given minimal input such as oxygen and nutrients) under fairly tight genetic control. An alternative term, maturation, suggests that the transformations of early life transcend hypertrophy and hyperplasia, yet still follow a preset programme. But this neglects to consider the environment's shaping role, which, in the nervous system at least, includes learning.

So development is not just more than growth — it is more than maturation, requiring constant negotiation with the environment. Sometimes this truth has led to a refusal to try to tease out the different roles of maturation and learning. In Jean Piaget's theory of mental development, for example, the contributions of learning and of a tacitly assumed preset programme are deliberately obscured. In another model, really a metaphor, maturation and learning are viewed as the warp and woof — one blue, one yellow — that give a swatch of cloth a green colour. The claim is that attempting to separate the two contributions destroys the unique product of their interaction.

Of course, a thicker, denser blue warp makes the cloth a bluer green. These features

of the warp, not to mention the design and technique of weaving, help to explain the outcome. In the 1950s, the prescient psychologist Anne Anastasi saw that the real question is not "which?" or "how much?"; but "how?" Advances in genetics and brain science now leave us in no doubt that we can answer all three. But how do we address the "how" question?

In embryology, development always entails interaction, although the interactions often take place inside the organism. In the classic account, the dimpling of the vertebrate eye from a blob-like to a cup-like shape — the formation of the retina — occurs in response to a chemical signal from the overlying ectoderm. Soon after, the lens is formed from ectoderm when the brand-new eye cup sends back another signal.

Later, as neurons form and migrate around the brain, they are attracted, re-routed and stopped by molecular signals. Many of these come from other cells that guide or challenge the migrators in a kind of immunochemistry. Recognition of cells and surfaces, and ultimately adhesion to them, determine the fates of neurons, and subsequently those of their extensions. These patterns become the wiring plan of the brain.

But this line of thinking comes up against Changeux's paradox: how do 30,000 human genes determine 10^{11} cells with 10^{15} connections? Obviously they can't do it in the same way that the roundworm's 18,000 genes govern its 959 cells. There are several solutions.

First, pioneer cells and axons pave the way for thousands or tens of thousands of others to track their guidance, offering lots of hook-ups for the price of one. Second, the mammalian brain forms many more cells and connections than it needs, subsequently pruning back around half of them. Some of this occurs through programmed cell death, but much depends on activity — meaning that spontaneous and reactive fetal movements shape the brain. Third, small groups of neurons may form under strict genetic control — creating small, deterministically wired systems similar to the roundworm's brain — and then compete for incoming stimulation and outgoing actions.

These processes have been called darwinian, but this is only a partial analogy. The cells of the embryo are genetically identical, and they produce no offspring, thus undermining two pillars of Darwin's theory — variation and inheritance. Still, the processes involve competition, which is resolved by environmental, adaptive selection. And the cells are not quite genetically identical — the same set of genes is always there, but only

Development

Development is not just more than growth — it is more than maturation, requiring constant negotiation with the environment.

some are switched on. Which switch on and which off in any given cell — and when, and how, and why — determine the cell's character and function. A main key to development is this on-off pattern, a pulsing, embryo-wide light show that turns genetic instructions into animals.

Elucidating the control of these switches — by signals inside the cell, beyond it, or even outside the body — is the main task of biology in the twenty-first century. And the switches are not flipped just in early life — genes that confer Huntington's and Alzheimer's diseases are switched on decades after the die is cast. But of course, in a complex animal, much is left to chance. Chaos in the formal sense — exquisite sensitivity to variations in starting conditions — cumulatively amplifies small differences. This embryonic butterfly effect gives identical twins different brains within weeks of conception. Such unpredictable paths help to explain why twins differ before we even consider their environmental influences.

Less certain is the role of emergence in development, but if self-organizing processes occur in non-living solutions, why not in a minuscule protoplasmic pool or an early, inchoate blob of cells? In computer models of embryos, self-organization looks to be adequate for certain tasks. We need to learn more about these less deterministic routes to life's complexity.

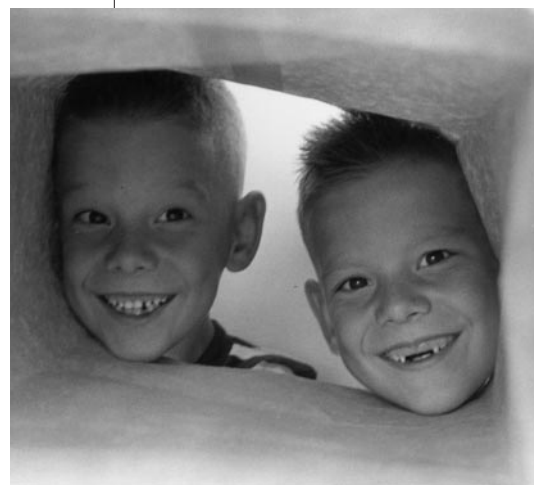
One thing is certain. The sequencing of the genome will soon look like the easiest thing that biologists ever did. And what sequencers euphemistically call "annotation" and the rest of us call development — what the genes actually do — constitutes the real code of living systems. To crack that code will take centuries, but getting there will be more than half the fun. ■

Melvin Konner teaches at Emory University, Atlanta, Georgia, USA. He is the author of the completely revised edition of *The Tangled Wing: Biological Constraints on the Human Spirit*.

FURTHER READING

Anastasi, A. *Psychol. Rev.* **65**, 197–208 (1958).
Changeux, J.-P. *Neuronal Man: The Biology of Mind* (trans. Garey, L.; Princeton Univ. Press, 1997).
Edelman, G. M. *Neural Darwinism: The Theory of Neuronal Group Selection* (Basic, New York, 1987).
Wolpert, L. *The Triumph of the Embryo* (Oxford Univ. Press, 1991).

GETTY IMAGES



Two of a kind? From before birth, chance and the environment conspire to make twins differ.

Survival of the entangled

William Barnes

At the quantum level, common sense is often violated — for example, by pairs of entangled photons in which each seems to ‘know’ about the state of the other. Entanglement may be more robust than had been thought.

Two particles are just two particles, aren't they? Well, no, actually — not always. In the weird world of quantum mechanics, particles can lose their separate identities to become just one ‘entangled’ object. Entanglement not only represents one of the biggest differences between the quantum and the classical, it also provides the key to exploiting quantum mechanics to build new technologies such as quantum computers. But entanglement is fragile: disturb the system and the entanglement is easily lost. However, Altewischer *et al.*¹ report, on page 304 of this issue, that entanglement might be better at surviving than we had suspected. Their findings could even point to a new way of manipulating entanglement for quantum technology.

Altewischer *et al.* launched photons of one energy (one wavelength) into a crystal whose particular properties cause each photon to be transformed into two new photons, a process called down-conversion. By the principle of conservation of energy, both down-converted photons have half the energy (twice the wavelength) of the original photon. Not only is energy conserved, so too is a more exotic quantity called spin. The consequence of spin conservation is that the polarizations of the two down-converted photons are always orthogonal (at right angles) to each other. For example, if we measure one photon to be linearly polarized in the vertical direction, we always find the other photon to be linearly polarized in the horizontal direction.

Now for some quantum weirdness. No matter which axis is chosen for the measurement of the polarization of one of the photons, that choice completely determines what happens when we measure the polarization state of the other photon². It is as though the two photons know about each other instantaneously — in defiance of special relativity, which states that no signal could pass between the photons faster than the speed of light. This strange effect is what we call entanglement. Even stranger, this mutual knowledge holds even if the polarization axis for our measurement is not chosen until after the photons have flown a considerable distance apart.

The key contribution of Altewischer *et al.*¹ is to show that entanglement can survive even when one (or both) of the entangled photons is converted into a ‘surface plasmon’

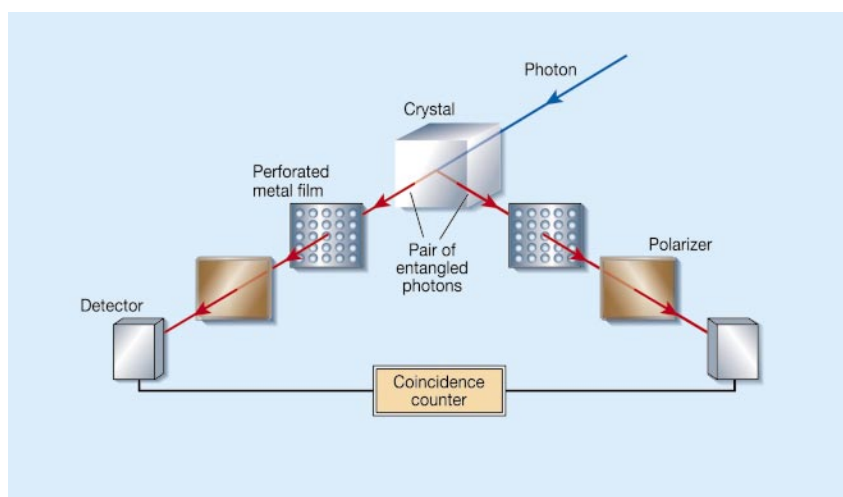


Figure 1 Quantum entanglement. A blue photon undergoes ‘down-conversion’ inside a crystal to form two lower-energy red photons that are entangled — that is, the measured polarization of one uniquely determines the polarization of the other. Polarizers placed in the path of these photons are set so that they pass orthogonal polarizations. The coincident signals in two separate detectors verify this weird quantum effect. Altewischer *et al.*¹ placed metal films, perforated with arrays of holes smaller than the photon wavelength, in the path of the photons. The entangled photons are each transformed into electron vibrations, called surface plasmons, on the metal films. These surface plasmons are short-lived, and the photons are re-emitted, still in their entangled state. The fact that quantum entanglement can withstand this conversion raises the possibility of controlling entangled photons through the manipulation of surface plasmons.

and then back into a photon. Surface plasmons are oscillating electromagnetic fields, strongly localized at the surface of a metal and associated with the collective motion of a large number of electrons. In the Altewischer experiment (Fig. 1), the down-converted photons were fired through metal films perforated by an array of holes smaller than the wavelength of the photons. Such metal films transmit photons surprisingly well³: incident photons scatter off the periodic arrangement of holes and are converted into a surface-plasmon mode on the metal surface; the electromagnetic field associated with the surface plasmon tunnels through the holes and is in turn scattered by the periodicity of the structure, finally being converted back to a photon. Altewischer and colleagues found that, although many photons are lost as a result of absorption in the metal film, some survive and, amazingly, are still entangled.

Given the collective nature of the surface plasmon, which involves a large number of electrons, it seems remarkable that entanglement should survive. It is perhaps all the

more surprising because surface-plasmon modes are short-lived; typically they are lost to absorption in the metal in just a few femtoseconds ($1 \text{ fs} = 10^{-15} \text{ s}$). But perhaps we shouldn't be so surprised. We would expect entanglement to survive if we simply used a metallic mirror to reflect an entangled photon, and we would still be making use of the collective motion of many electrons to provide the reflection. In both the surface-plasmon and simple reflection processes, we rely on the electron–electron scattering rate being low enough to allow the electron motion to remain coherent.

The advantage of using surface plasmons is that they are well-defined modes, and, by building suitable nanoscale structures, we can control them on length scales small enough to ensure that the electron motion remains coherent⁴. In fact, we can control surface plasmons well enough to achieve highly concentrated optical fields in volumes smaller than the wavelength, thus allowing us to control the flow of optical energy on short length scales. Altewischer *et al.*¹ have demonstrated that entanglement

survives when surface plasmons are used. The challenge now is to see whether we can exploit the combination of nanotechnology and surface plasmons to manipulate entanglement for the benefit of emerging quantum technologies.

Despite our increasing ability to control entanglement, it still seems totally at odds with common-sense thinking — how can the photons know about each other when special relativity precludes signals getting from one to the other quickly enough to influence the outcome of our measurements? In finding entanglement difficult to comprehend we are in good company: Einstein and others were so disturbed by the concept that they questioned the validity

of quantum mechanics⁵. Such conceptual problems arise because of our natural tendency always to think of photons as separate, well-defined objects: nature is not subject to such limitations. ■

William Barnes is in the School of Physics, University of Exeter, Stocker Road, Exeter EX4 4QL, UK.
e-mail: w.l.barnes@exeter.ac.uk

1. Altevischer, E., van Exter, M. P. & Woerdman, J. P. *Nature* 418, 304–306 (2002).
2. Aspect, A., Grangier, P. & Roger, G. *Phys. Rev. Lett.* 47, 460–463 (1981).
3. Ebbesen, T. W., Lezec, H. J., Ghaemi, H. F., Thio, T. & Wolff, P. A. *Nature* 391, 667–669 (1998).
4. Krenn, J. R. *et al.* *Phys. Rev. Lett.* 82, 2590–2593 (1999).
5. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* 47, 777–780 (1935).

Developmental biology

Decisions, decisions!

Brigid Hogan

Early embryo cells can develop either into specialized body cells or into precursors of eggs or sperm. It is not understood how this crucial decision is made in mammals, but new work brings us closer to the answer.

One of the most contentious issues in biology today is whether a stem cell from an adult tissue, committed to generate a few specialized cell types, can be 'reprogrammed' to produce many more. Pushing the envelope, some have even suggested that an adult cell might be persuaded to give rise to eggs and sperm¹ — our precious germ cells, which together can generate a complete organism. But there are still huge lacunae in our knowledge of the processes we are trying to reverse. How do cells of the early embryo normally become limited in their developmental potential? And how are a privileged few embryonic cells chosen to become germ cells? On page 293 of this issue, Saitou and colleagues² describe a bold approach to these problems. The authors monitored gene activity in individual mouse embryonic cells at precisely the time when a group decision was being reached as to which would become body cells and which the germ cells.

In rapidly developing organisms such as worms and flies, germ-cell fate is decided very autocratically. Material known as germ plasm, which dictates germ-cell development, is deposited in the egg before fertilization. It is then parcelled out to specific cells early in development, sealing their fate^{3,4}. By contrast, in mammalian embryos, germ-cell status is acquired in a democratic way, as a result of interactions between neighbouring cells. In mice, these interactions are initiated when there are only three cell layers in the embryo (Fig. 1). The inner layer (the epiblast) eventually gives rise to all the cells of the fetus, including the germ cells,

while the outer layers (one of which is the 'extra-embryonic ectoderm') are supportive tissues that nevertheless send important early patterning signals to the epiblast.

Initially, the epiblast is a cup-shaped sheet, but during the process known as gastrulation cells move towards one side, drop

out of the layer and give rise to a new cell population, the mesoderm. Most mesoderm cells move into the developing embryo, while the remainder contribute to extra-embryonic mesodermal support tissues — the amnion and allantois. Early germ cells (or 'primordial germ cells', PGCs) are first identified around this time as a cluster of about 40 to 50 cells expressing high levels of the enzyme alkaline phosphatase (encoded by the *Tnap* gene)^{3,5}. This cluster lies in a special niche in the no-man's land between the embryonic and extra-embryonic mesoderm (Fig. 1).

Where do these PGCs come from? Cell-lineage studies show that they originate in the rim of the epiblast cup, right next to the extra-embryonic ectoderm⁵. Importantly, the fate of cells in this region is not yet sealed. Rather, it appears that the selection of future PGCs is made in two steps. First, in response to signals from the extra-embryonic ectoderm, epiblast cells in the rim are programmed to become common precursors of extra-embryonic mesoderm and PGCs. A second signal then imposes germ-cell status on a few of these cells; the others differentiate into extra-embryonic mesoderm.

Experiments have identified growth factors called bone morphogenetic proteins (Bmp4 and Bmp8b) as components of the first signal^{6,7}. To learn more about the second decision-making step, Saitou *et al.*² decided to investigate which genes are active in individual cells in the region where the PGCs expressing high levels of *Tnap* will appear.

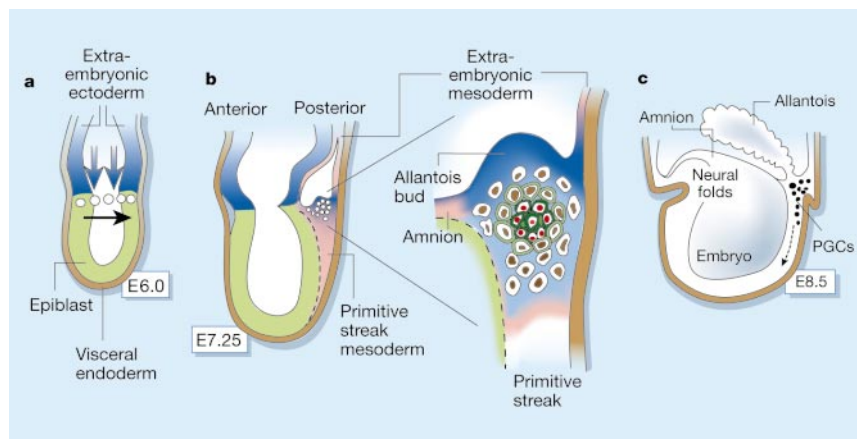


Figure 1 Model for the development of primordial germ cells (PGCs) in mouse embryos, according to Saitou *et al.*². a, Six days after fertilization (E6.0) the embryo consists of three layers. The epiblast will produce the fetus; the extra-embryonic ectoderm and the visceral endoderm are supportive tissues. Signals from the extra-embryonic ectoderm, including certain Bmp proteins (blue), induce neighbouring epiblast cells (open circles) to become common precursors of extra-embryonic mesoderm and PGCs. During gastrulation these cells move (black arrow) towards the posterior and drop out of the epiblast. Saitou *et al.* predict that, as the cells migrate, they adhere to one another through the membrane protein fragilis. b, This clustering becomes more pronounced by E7.5. Cells in the centre, which express the highest levels of the *Tnap* and *fragilis* proteins (dark green), are induced to express the nuclear protein *Pgc7/stella* (red) and will become PGCs. Cells at the periphery and outside the cluster express *Hoxb1* (brown). The Bmp present in this region (lighter blue) promotes the survival of PGCs¹⁴. Precisely what drives decision-making in the cluster is not known. c, Once PGCs are specified, they migrate to the future gonads (dashed arrow).

They isolated about 80 of these cells and made copies (cDNAs) of all of the messenger RNAs expressed in each of them, keeping them as separate pools. As the cells all look alike at this time, the authors had to sort the cDNA pools to select those likely to be derived from future PGCs rather than their neighbours (which are being slated to form mesoderm). To do this, they selected pools containing many copies of the *Tnap* cDNA and rejected any that were rich in cDNAs for *Bmp4* and *Hoxb1* — genes known from other studies to be expressed predominantly by mesoderm cells.

Saitou *et al.* then identified novel genes that are preferentially represented at high levels in a selected *Tnap*-expressing pool. One of these genes, named *fragilis*, encodes a protein that the authors predict is localized in the outer cell membrane. On the basis of studies with related proteins⁸, it is likely that cells with the *fragilis* protein on their surface cluster together.

Saitou *et al.* show that expression of *fragilis* first increases in the rim of the epiblast cup, suggesting that one of its roles is to keep together those cells that are predisposed to become PGCs, as they move out of the epiblast during gastrulation. The discovery of *fragilis* at this stage also raises other intriguing questions. Immunological studies have shown that the expression of *fragilis*-related proteins can be induced by interferons (a large family of proteins that normally suppress the growth of cells of the immune system). This hints that the interferons might have roles in PGC development. Unfortunately, this will be difficult to test, as there are many genes in the mouse that encode interferons, their receptors and the molecules they signal to^{9–12}.

A second gene identified in Saitou *et al.*'s screen encodes a putative DNA-binding protein named *stella*, which probably acts in the cell nucleus to control gene expression. This protein had already been identified recently in PGCs by another group¹³, and called *Pgc7*. Interestingly, both groups^{2,13} show that *Pgc7/stella* is not unique to PGCs but is also expressed in fertilized eggs, in the cells of early mouse embryos and in embryonic stem cells — all of which have a very high developmental potential. This suggests that one role for the protein is to promote the activity of genes that keep cells undifferentiated, or to repress those that drive differentiation forward.

How do these results tie together? Saitou *et al.*² propose an elegant model in which cells at the centre of the community of precursors, which express the highest levels of *fragilis*, are the ones selected to become PGCs and to express high levels of *stella*. Precisely how this works can now be tested experimentally. The results are likely to provide new insights into the way in which the developmental potential of embryonic cells

normally becomes restricted, and how germ cells avoid this fate. This information should, in turn, help us understand how such decisions could be artificially reversed.

Brigid Hogan is at the Howard Hughes Medical Institute, Department of Cell and Developmental Biology, Vanderbilt University Medical School, C-2310 Medical Center North, Nashville, Tennessee 37232-2175, USA.

e-mail: brigid.hogan@vanderbilt.edu

1. Zhao, G. Q. & Garbers, D. L. *Dev. Cell* **2**, 537–547 (2002).
2. Saitou, M., Barton, S. C. & Surani, M. A. *Nature* **418**, 293–300 (2002).
3. Wylie, C. *Cell* **96**, 165–174 (1999).
4. Seydoux, G. & Strome, S. *Development* **126**, 3275–3283 (1999).

5. Lawson, K. A. & Hage, W. J. in *Ciba Foundation Symposium Germline Development* (eds Marsh, J. & Goode, J.) 68–91 (Wiley, Chichester, 1994).
6. Lawson, K. A. *et al. Genes Dev.* **13**, 424–436 (1999).
7. Ying, Y., Liu, X.-M., Marble, A., Lawson, K. A. & Zhao, G.-Q. *Mol. Endocrinol.* **14**, 1053–1063 (2000).
8. Evans, S. S., Collea, R. P., Leasure, J. A. & Lee, D. B. *J. Immunol.* **150**, 736–747 (1993).
9. Aaronson, D. S. & Horvath, C. M. *Science* **296**, 1653–1655 (2002).
10. Niwa, H., Burdon, T., Chambers, I. & Smith, A. *Genes Dev.* **12**, 2048–2060 (1998).
11. Mitani, Y. *et al. Genes Cells* **6**, 631–640 (2001).
12. Truchet, S., Wietzerbin, J. & Debey, P. *Mol. Reprod. Dev.* **60**, 319–330 (2001).
13. Sato, M. *et al. Mech. Dev.* **113**, 91–94 (2002).
14. Fujiwara, T., Dunn, N. R. & Hogan, B. L. *Proc. Natl Acad. Sci. USA* **98**, 13739–13744 (2001).

Population genetics

Malaria variorum

Andrew G. Clark

A genetic study of the malaria parasite finds that this species is unexpectedly diverse. A second study shows multiple independent origins of mutations in one parasite gene that confer resistance to a widely used drug.

Understanding genetic variation in the malarial parasite, *Plasmodium falciparum*, is of major importance to public health, especially as we contemplate widespread programmes of vaccination. If a vaccine controls only part of the *P. falciparum* population, then it might alter the genetic composition of the surviving parasites. The frightening prospect is that such an imperfect vaccine might select for variants of *P. falciparum* that are more virulent, potentially undoing any progress towards mitigating this scourge¹. A thorough knowledge of the genetic diversity of the parasite population is the first defence against this possible disaster — hence the importance of papers by Mu *et al.*² and Wootton *et al.*³ on pages 323 and 320 of this issue.

There is considerable disagreement in the scientific literature regarding the level of variability of existing *P. falciparum* isolates and, by inference, their depth of common ancestry (that is, the time at which they started to diverge). All agree that there is abundant variation in microsatellites — simple repetitive DNA sequences. And it cannot be disputed that different *P. falciparum* isolates vary in their ability to resist drugs and in molecules that trigger immune responses in humans. But studies of microsatellites and protein differences might not be optimal when making inferences about the long-term population size and time of common ancestry of different isolates. ‘Silent’ DNA changes, which do not alter encoded amino-acid sequences, provide a picture that is more clock-like and less distorted by natural selection. Some have argued that there is an exceptionally low level of silent variation in *P. falciparum* isolates, and a correspondingly

recent time of common ancestry^{4,5}, around 3,000–10,000 years ago. Others — including Mu *et al.* — have suggested a date of 50,000 years ago or more^{2,6}.

Why do these estimates vary so wildly? Part of the answer is that the studies differed in several significant ways, including the criteria by which the isolates were identified, which genes were examined, and how the data were analysed. Rich *et al.*⁴ randomly sampled the protein-coding regions of 9 genes in 35 *P. falciparum* isolates, and found differences (none of which were in fact silent) at 51 positions in the sequences. The authors carefully calibrated the rate of evolution of the sequences, comparing them with data from rodent and chimpanzee *Plasmodium* parasites, yielding an expected time of common ancestry of 6,000–12,000 years ago. Volkman *et al.*⁵, meanwhile, sequenced 25 introns (non-protein-coding sequences in genes) from 8 *P. falciparum* isolates. They found 8 varying bases; 5 were either in or near microsatellites and so were excluded from further analysis, which gave a date of common ancestry of 3,200–7,700 years ago. These studies are consistent with the idea that malaria became a widespread human pathogen only after the spread of agricultural practices that expanded the range of the malaria-transmitting mosquito⁷. Another implication is that the estimated low level of variation should make control easier.

Hughes and Verra⁶, however, calculated a much higher degree of variation, consistent with a parasite population size of the order of 100,000 for the past 300,000–400,000 years. Extracting a larger data set from public databases, these authors had identified 23 genes that satisfied their criteria for ‘neutral’



100 YEARS AGO

Symbol for Partial Differentiation. "The insufficiency of this notation is not forgotten, however, although its advantages over the different devices of Euler and Lagrange are recognised...". I am glad to think that a pure mathematician sees the difficulty met with by users of mathematics. I wish that men who write to me privately would publish their remarks. One correspondent says: "I think 'the mathematicians' made a rather stupid blunder when they introduced ∂ for partial differentiation. This way: nearly all differential coefficients are partial; even a complete one (assumed complete) may become partial by extension of the field of operation. So an old investigation of Kelvin's, for example, using d throughout, is, by 'the mathematicians', replaced by the same using ∂ throughout, except one or two here and there! What is the use? It gives a lot of trouble and as the printers haven't always ∂ 's, or proper sized ∂ 's, it makes bad work. It should have been ∂ itself that was introduced for the exceptional use, thus making next to no alteration in the classical investigations." These are indeed my own views, but as my pupils go forward to University examinations I advise them to adopt the fashion which is likely to please the examiners.

From *Nature* 17 July 1902.

50 YEARS AGO

Messrs. Blundell Rules, Ltd., Chaul End Lane, Luton, Beds, have issued a small handbook of 22 pages (price 6d.) describing their range of slide rules, and giving examples of the uses of the various patterns. Thus, in addition to the usual rules in various styles graduated for simple mathematical operations, there are others with log-log scales, commercial calculators, electricians' scales with appropriate quantities, and a timber calculator. Most of these scales are made entirely of a white plastic material, based on vinyl chloride-vinylidene chloride copolymers, with a 'satin' finish. It is claimed to be resistant to water, most chemicals and fungi. It is strong, and the scales are sharp and clear; in some patterns the tension of the slide can be adjusted. Model A.G.1 (£1 13s.), a simple general-purpose rule, has the usual A , B , C , D scales, and also scales for reciprocals, cubes and mantissæ; additional red lines on the cursor provide for finding areas of circles of given diameter.

From *Nature* 19 July 1952.

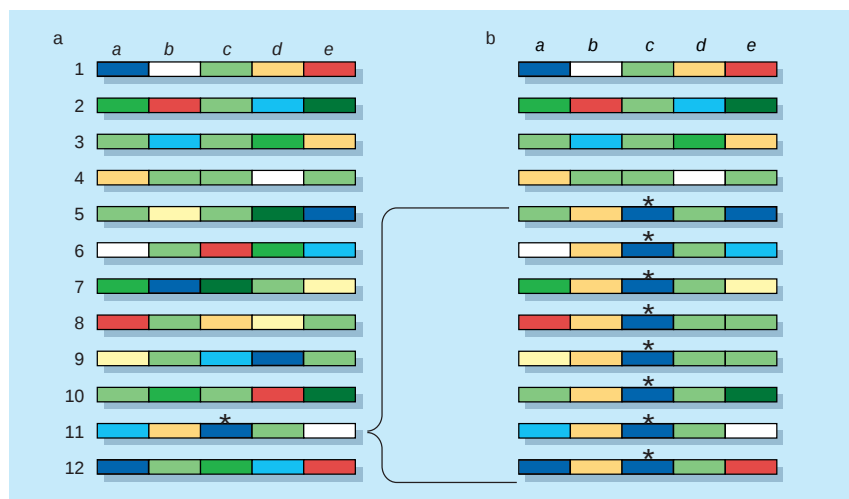


Figure 1 Genetic hitchhiking. a, A single chromosome with 5 genes ($a-e$), from 12 isolates, is shown. Each gene has considerable variation (it comes in different alleles, depicted by different colours). When a favourable mutation occurs (asterisk), it increases in frequency (b), taking not only the blue version of gene c to high frequency, but also the yellow form of gene b and the green form of gene d . These three genes end up losing most of their allelic variation after such 'hitchhiking'. Genes farther away from the favoured gene c have undergone exchanges with other chromosomes (recombination), so the loss of variation decreases with greater distance from the selected gene.

evolution — there was no strong indication that natural selection had distorted the pattern of variation. They discovered 159 variable positions, including 11 silent changes, within these sequences. One weakness, not unique to this study, was in accepting the public databases as error-free, so any sequencing mistakes would have been incorporated into their model as variation. But even if error rates were at the high end of credibility, the level of variation in these data still exceeds that of the other studies^{4,5}.

This sets the stage for the approach of Mu *et al.*², who re-sequenced portions of 204 genes in the largest survey so far of variation on chromosome 3 of *P. falciparum*. The authors obtained more than 218,000 base pairs of coding and non-coding sequence from 5 widely divergent isolates, and found 238 single changes (62 of which were silent and 31 of which were in non-coding sequence) and 165 microsatellite variants. From these data Mu *et al.* estimate a relatively high level of variation at silent sites, consistent with a time of common ancestry of 150,000 years ago.

All these details are important because there have been several competing explanations for the many different estimates of *P. falciparum* ancestry. One proposed explanation is that *P. falciparum* is a recently evolved species, with mutation rates that vary wildly along its chromosomes, giving different pictures depending on the genes studied. Another is that *P. falciparum* is ancient but in the recent past suffered a massive restriction ('bottleneck') in the number of surviving lineages and consequently lost most variation — leading to the relatively young date of common ancestry calculated in studies that,

by chance, failed to sample any divergent but rare alleles. A third is that natural selection favoured novel mutations, such as those for drug resistance, and that as these mutations increased in frequency, they dragged linked genes (in a process known as 'hitchhiking'; Fig. 1) with them. This too would create an impression of little variation in some genes. Although it might not settle the question, Mu and colleagues' study² of DNA sequence variation is far larger than its predecessors, and provides strong evidence that *P. falciparum* is diverse enough to raise a greater challenge for public-health measures.

In a separate analysis³ the same group provide a beautiful example of genetic hitchhiking. Wootton *et al.*³ were looking at variation in the ability of *P. falciparum* to resist the widely used antimalarial drug chloroquine. They studied microsatellites in 87 worldwide isolates of the parasite, and found that a narrow region of chromosome 7 — including the *pfprt* chloroquine-resistance gene and flanking sequences — shows little variability among chloroquine-resistant isolates but is more diverse among sensitive isolates. Further analysis supported the idea that there are at least four independent forms of *pfprt*, which are rapidly becoming more common in different parts of the world. The inference that selective pressures (in this case, the use of chloroquine) cause such localized increases in gene frequency is receiving attention elsewhere⁶, but a technical advantage of *Plasmodium* is that it has just one copy, not two, of each chromosome. The results show that complex combinations of mutations arise readily. So, in principle, *P. falciparum* might rapidly develop resistance to multiple drugs. The genes also

seem to have moved across continents with frightening speed, implying that there would be little time to contain the spread of new resistance genes.

How can this information help us to control malaria? First, although it is tempting to think that we now know the full extent of *P. falciparum* diversity, further surveys are still needed. A present-day snapshot of variation in many genes among more isolates, spanning broad geographical areas, will be important in determining the genetic impact of a vaccination or expanded drug programme. Had thorough genetic monitoring been in place, we might have detected the genetic variants causing the spread of chloroquine resistance early enough to take steps to prolong the usefulness of the drug. Second, we need a systematic collection of *Plasmodium* isolates to be shared among

laboratories, common resources for identifying and confirming sequence variations on a large scale, and a common database of information. We should then rapidly reach consensus on the genetic variability and ancestral demographic history of this devastating parasite.

Andrew G. Clark is in the Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA.

e-mail: andyclark@cornell.edu

1. Gandon, S., Mackinnon, M. J., Nee, S. & Read, A. F. *Nature* **414**, 751–756 (2001).
2. Mu, J. *et al.* *Nature* **418**, 323–326 (2002).
3. Wootton, J. C. *et al.* *Nature* **418**, 320–323 (2002).
4. Rich, S. M., Licht, M. C., Hudson, R. R. & Ayala, F. J. *Proc. Natl Acad. Sci. USA* **95**, 4425–4430 (1998).
5. Volkman, S. K. *et al.* *Science* **293**, 482–484 (2001).
6. Hughes, A. L. & Verra, F. *Proc. R. Soc. Lond. B* **268**, 1855–1860 (2001).
7. Coluzzi, M. *Parassitologia* **41**, 277–283 (1999).
8. Kim, Y. & Stephan, W. *Genetics* **160**, 765–777 (2002).

Materials science

Plastic parameter

Jonathan A. Zimmerman

Materials may deform permanently under pressure, but it's difficult to guess exactly what will happen to their structure at the atomic level. A new model, and a powerful parameter, might allow firm predictions to be made.

A child tries to dish out ice cream using a metal spoon, but the neck of the spoon bends back permanently with the effort applied. A construction worker attempts to drive a nail into a plank, only to hit it slightly off-centre and end up with a useless bit of metal that must be extracted from the wood. These examples show that 'plasticity', the irreversible deformation of a material, affects people's lives on a daily basis. But how plastic deformation occurs at the atomic scale is not fully understood. A new model, presented by Li *et al.*¹ on page 307 of this issue, takes us further into the details of this complicated process.

The term 'plasticity' is most commonly used for metals and other materials that have an underlying crystal structure, where the constituent atoms are arranged into a lattice with a repeatable pattern. When a line defect, or dislocation, is created in the structure, it moves through the material, resulting in a permanent change in its shape². This movement is called 'slip', as it refers to the slipping of adjacent atomic planes past one another. There has been much fundamental research into what triggers, or nucleates, a dislocation and how plasticity develops in crystals. As the characteristic dimension of dislocations — the Burgers vector — is of the order of the spacing between atoms, these dislocations are studied using nanoscale experimental techniques and modelling at the level of individual atoms — atomistic simulation.

However, it is difficult to link the mechanisms at work in simulations to the observations made in larger-scale physical experiments. Li *et al.*¹ impressively combine aspects of atomistic simulation with finite-element analysis treating the crystal as a continuum, to predict the onset of dislocation during nanoindentation — when the crystal is indented to nanometre-scale depths. They

also quantify the features of the dislocations created and predict how the crystal microstructure will change with further indentation. Their model captures many key features observed in experiments.

A concept emphasized in their model is the stability parameter. This is an extremely useful tool for analysing materials simulations, as it identifies the location within the material at which defect nucleation is occurring and the characteristics of the defect being created. For example, indentation of a metal commonly results in the creation of a dislocation. Li and colleagues' parameter not only indicates at which atom the dislocation will originate, but also has information on the Burgers vector of the dislocation and the crystallographic plane on which slip occurs. Alternatively, a brittle material undergoing tensile loading may ultimately fail because of the creation of a 'microcrack'. Presumably, their parameter would again indicate the atom at which the flaw is spawned as well as providing quantitative information, such as the orientation of the cleavage plane.

These situations are straightforward, but there are some materials that exhibit both brittle and ductile behaviour, the latter occurring through deformation by the creation and propagation of dislocations. The transition between ductile and brittle behaviour is an important area of research that has received much attention in recent years. Li *et al.*'s new parameter will allow a thorough investigation, through simulation, of the conditions that lead to one response or the other.

A typical case is shown in Fig. 1, where the mechanical loading at the tip of some pre-existing crack in a crystal can result in either the atomic planes splitting apart or the

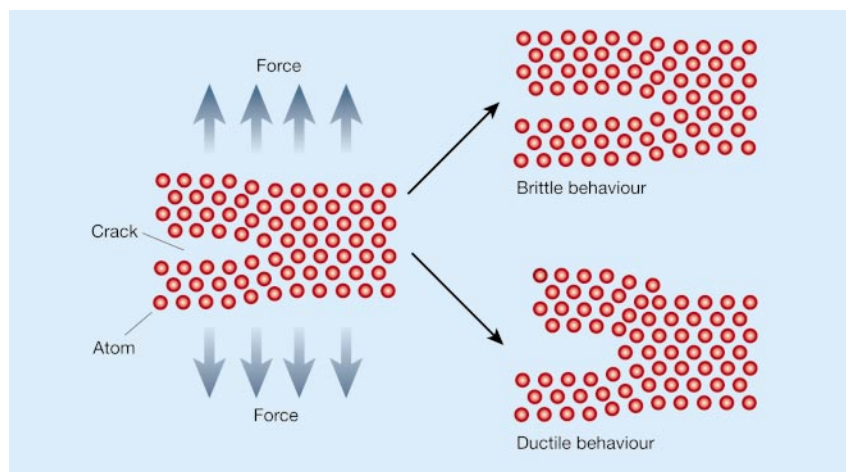


Figure 1 Cracking up. When force is applied to a sharp, atom-wide crack in a material, brittle or ductile behaviour may ensue. If the response is brittle behaviour (upper right), the atomic planes cleave as the crack propagates. With a ductile response (lower right), the tip of the crack becomes blunted as dislocations spread through the crystal structure. Predicting whether a material will respond in a brittle or a ductile fashion to an applied load is complicated; many factors are involved. But Li *et al.*¹ have devised a 'stability parameter' that effectively models the formation and propagation of dislocations through a material.

emission and propagation of dislocations, depending on how the load is applied. Analytical^{3,4}, computational^{5,6} and experimental^{7,8} studies have been carried out in an attempt to pin down the relationships between applied thermo-mechanical loading and the way in which a material will react. A combination of factors are usually involved: temperature⁸; motion or dynamics of existing defects^{6,7}; the specific elements and alloys that make up the crystal⁵; the orientation of the crystal lattice with respect to the load; the rate at which the load is applied; existing microstructural barriers such as crystal-grain boundaries; and properties of the material that determine its resistance to ductile and brittle deformation.

This last factor was well summarized by Rice³, who compared the energetic 'cost' of creating new surfaces within the crystal (the surface energy) with the energy barrier that must be overcome to allow atomic planes to slip over one another (the unstable stacking fault energy). Rice used fracture mechanics to perform a straightforward analysis, and found that a critical ratio of these energies separates ductile from brittle behaviour for a particular crystal-lattice type and orientation.

Given the above list of factors, it is clearly difficult to predict whether a material will fail in a ductile or a brittle fashion. Materials scientists try to determine the competing mechanisms involved and to identify properties in the material that influence nucleation and propagation. Prevailing theories of failure mechanics use quantities that are related to the load-bearing ability of a material, such as yield and fracture stresses, whereas newer methods rely on evaluating the energetic barriers described above. Complex systems will require a combination of these criteria to adequately describe any irreversible deformation process. Although it has yet to be used for this, the model devised by Li *et al.*¹ has the potential to predict whether brittle or ductile behaviour will occur.

It is in the area of predictive modelling that the work of Li *et al.* will be immensely useful. The authors present a methodology for using their stability parameter in finite-element calculations — a simulation technique that models materials as a continuum, but uses the same interatomic potentials as in atomistic simulations. This enables systems with larger-than-atomic dimensions to be simulated, and can take into account microstructural features such as grain boundaries and different phases of the material. There is also the exciting possibility of refining the approach to operate effectively at finite temperatures and high loading rates. Li *et al.* have achieved a significant step towards understanding the defective world of materials in which we live. ■

Jonathan A. Zimmerman is at Sandia National

Laboratories, PO Box 969, Livermore, California 94551-0969, USA.

e-mail: jzimmer@sandia.gov

1. Li, J., Van Vliet, K. J., Zhu, T., Yip, S. & Suresh, S. *Nature* **418**, 307–310 (2002).
2. Askeland, D. R. *The Science and Engineering of Materials* (PWS-Kent, Boston, Massachusetts, 1994).

3. Rice, J. R. *J. Mech. Phys. Solids* **40**, 239–271 (1992).
4. Beltz, G. E., Lipkin, D. M. & Fischer, L. L. *Phys. Rev. Lett.* **82**, 4468–4471 (1999).
5. Farkas, D. *MRS Bull.* **25**, 35–38 (2000).
6. Abraham, F. F. *J. Mech. Phys. Solids* **49**, 2095–2111 (2001).
7. Wall, O. *Eng. Fract. Mech.* **69**, 835–849 (2002).
8. Šmida, T. & Božanský, J. *Mater. Sci. Eng. A* **323**, 21–26 (2002).

Cell biology

The extraordinary phagosome

Joel A. Swanson

The finding that a cellular compartment called the endoplasmic reticulum can merge with the cell's outer membrane is surprising. It would not have been predicted from our knowledge of cell organization.

Within our cells are many membrane-bounded compartments, each of which communicates with only a subset of other such 'organelles' — generally when small vesicles bud off from one and merge with another. The efficiency with which cells use their organelles to carry out complex tasks has led to the belief that movement through the system of organelles occurs in strictly defined directions. So, the endoplasmic reticulum (ER) — in which proteins destined for export from the cell are synthesized and modified — communicates primarily with the Golgi apparatus. This in turn interacts with a limited set of other organelles and, eventually, with the plasma membrane, defining an avenue of vesicular traffic for protein export. In the reverse direction, extracellular molecules internalized by endocytosis first enter endosomes and later reach lysosomes, where they are degraded (Fig. 1).

Although lateral communication between the organelles for export and import is well documented, their vectorial organization has been supported by the positions of ER and lysosomes at the ends of the two avenues. But faith in these traditional routes, already shaken by studies of phagosome biogenesis¹, is tested again by a report from Gagnon and colleagues² in *Cell*, which shows that proteins characteristic of the ER are present in phagosomes. This has implications for our understanding of cellular organization and of how microbes manipulate that organization.

Phagosomes are membrane-bounded organelles formed when one cell engulfs another, or some inanimate particle, by enclosing it in surface membrane. Early studies indicated that new phagosomes were made of plasma membrane³, and that the subsequent merger of phagosomes with lysosomes created a phagolysosome — a toxic environment for ingested microbes. By current thinking, phagolysosomes form by progressive interactions of phagosomes with endosomes and lysosomes⁴. Delivery to phagolysosomes is considered the end of the

road and a fate for ingested microbes to avoid, if possible.

So pathogens that live within host cells enter by mechanisms that often resemble phagocytosis, but then take different routes. An early challenge to the conventional view of organelle interactions came from studies of *Trypanosoma cruzi*, which enters host cells by stimulating the fusion of lysosomes directly with the plasma membrane¹. Other pathogens, such as *Legionella pneumophila* and *Brucella abortus*, are ingested by phagocytosis, inhibit the fusion of phagosomes with lysosomes, and then inhabit an ER-like compartment⁵.

And perhaps such seemingly extraordinary routes are not as uncommon as one might think. Gagnon *et al.*² have now found the ER acting out of order, fusing with the plasma membrane and apparently providing membrane for phagocytosis. Using a proteomics approach the authors first showed that phagosomes containing latex beads — purified from extracts of professional phagocytic cells known as macrophages — displayed several ER marker proteins. Electron microscopy then showed ER membranes in continuity with the plasma membrane near cell-bound particles. Moreover, early phagosomes contained patches of both ER and lysosomal membranes. Gagnon *et al.* also found that newly formed phagosomes were enriched with ER proteins, and, as phagosomes aged inside macrophages, the levels of ER markers diminished and the number of lysosomal markers increased. Strangely, the levels of ER markers in phagosomes increased again later, indicating that phagosomes continued to fuse with ER membranes long after phagocytosis.

Although further evidence will be needed before we can say that the ER is essential to phagocytosis, it is reasonable to conclude that the ER supplies membrane to some phagosomes. Phagocytosis has been shown to involve fusion of intracellular organelles with the plasma membrane⁶; endosomes have been considered the most likely organelles, but lysosomes and the ER can

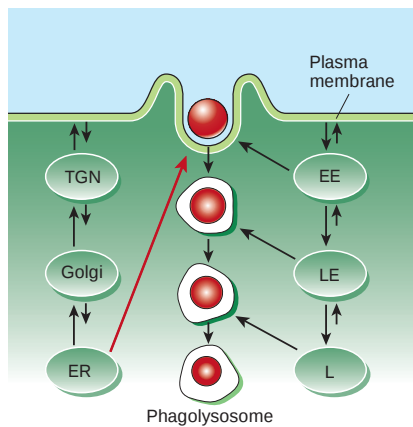


Figure 1 Intracellular organelles that interact with phagosomes — membrane-bounded compartments for the uptake of certain extracellular substances. Organelles involved in protein synthesis and secretion, including the endoplasmic reticulum (ER), Golgi apparatus and trans-Golgi network (TGN), have not been thought to interact significantly with phagosomes. Rather, phagosome formation at the plasma membrane leads to progressive interactions with endocytic compartments — early endosomes (EE), late endosomes (LE) and lysosomes (L) — ultimately forming a phagolysosome. But the paper by Gagnon *et al.*² shows that ER markers are present in phagosomes at the earliest stages of their formation (red arrow).

provide membrane as well. It is not clear why the delivery of ER to phagosomes was not discovered in earlier electron-microscopic studies, especially given that Gagnon *et al.* found ER markers in several different classes of phagosome, including those that take up the microbes *Leishmania donovani* and *Salmonella typhimurium*. But perhaps ER recruitment is not needed for all types of phagocytosis in macrophages.

Whatever the reason, these results² clearly suggest new lines of organelle communication, simplifying some previously complex models. For example, there are implications for how proteins are processed by 'antigen-presenting' cells and displayed on the cell surface for detection by the immune system. Molecules from intracellular pathogens are generally loaded onto so-called class I major histocompatibility complex proteins in the ER. Sometimes proteins from outside cells can also be loaded via this ER pathway, and some circuitous routes have been proposed for how those proteins reach the ER⁷. The results of Gagnon *et al.* suggest a more direct route. Moreover, if the ER also interacts with the plasma membrane in the absence of phagocytosis — a process that may be more difficult to detect — it could provide a short cut for some bacterial toxins to travel from the cell surface to the ER. Until now, these toxins had been thought to reach the ER by travelling 'backwards' through the Golgi apparatus⁸.

The new findings also suggest that the exotic routes of some intracellular pathogens are not so odd after all. Many pathogens define their compartmental home even before they separate from the plasma membrane, rather than remodelling a conventional phagosome after entry⁹. For example, *Legionella pneumophila* is thought to recruit ER to phagosomes, apparently by hijacking a normal cellular pathway for organelle degradation called autophagy⁵. But if further studies show that a variety of phagosomes contain ER proteins from the start, then the mechanism that creates the *Legionella* compartment can be viewed as a variation on

normal phagocytosis. That said, there probably isn't a 'normal' route for a phagosome. Gagnon *et al.* find that the ER merges with phagosomes in macrophages but not in neutrophils — another kind of professional phagocyte. Apparently, even phagosomes containing inert latex beads follow different patterns of maturation after phagocytosis.

Finally, the extraordinary variation among phagosomes may also hint at greater diversity among other organelles. With the increased sensitivity of proteomic methods, we may learn that organelle marker proteins

are not as specific as previously thought, and that the boundaries that distinguish organelles are not as sharp. Vectorial movements through organelles might be controlled at the level of individual molecules or patches of membrane, rather than by constrained interactions between discrete organelles. Phagosomes might then acquire unusual combinations of markers through the way in which the particles they contain affect marker dynamics. The resulting variety in marker proteins could provide a way for cells to sense the contents of their phagosomes. ■

Joel A. Swanson is in the Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109-0620, USA.
e-mail: jswan@umich.edu

1. Andrews, N. W. *Trends Cell Biol.* 5, 133–137 (1995).
2. Gagnon, E. *et al.* *Cell* 110, 119–131 (2002).
3. Muller, W. A., Steinman, R. M. & Cohn, Z. A. *J. Cell Biol.* 86, 292–303 (1980).
4. Desjardins, M., Huber, L. A., Parton, R. G. & Griffiths, G. *J. Cell Biol.* 124, 677–688 (1994).
5. Dorn, B. R., Dunn, W. A. & Progulske-Fox, A. *Cell. Microbiol.* 4, 1–10 (2002).
6. Bajno, L. *et al.* *J. Cell Biol.* 149, 697–705 (2000).
7. Watts, C. & Amigorena, S. *Semin. Immunol.* 13, 373–379 (2001).
8. Sandvig, K. & van Deurs, B. *EMBO J.* 19, 5943–5950 (2000).
9. Amer, A. O. & Swanson, M. S. *Curr. Opin. Microbiol.* 5, 56–61 (2002).

Longevity

Don't hold your breath

Siu Sylvia Lee and Gary Ruvkun

In some organisms a reduced-calorie diet increases lifespan. Conventional thinking about the mechanism involved now comes under question from the results of experiments with yeast.

More than sixty years ago a report appeared demonstrating that rats fed a diet containing 30–50% fewer calories live for four years instead of the normal three¹. 'Calorie restriction' (CR) has since been shown to extend the lifespan of species ranging from unicellular yeast, to worms and flies, and certain mammals². As they describe on page 344 of this issue³, Lin *et al.* have discovered that CR increases lifespan in yeast by turning up the rate of respiration by a factor of three. This finding challenges the traditional view that CR extends lifespan by decreasing metabolism and the associated production of damaging free-radical forms of oxygen. These free radicals are a by-product of the electron-transport chain by which mitochondria — subcellular organelles — generate energy during respiration.

It is not yet known if CR in humans increases lifespan, but experiments under way in primates suggest that it does⁴. Prompted by the continuing stream of press reports about the rodent experiments,

a variety of optimists, hucksters and fanatics have started to promote CR and to practise it with the aim of living longer themselves — or at least transferring wealth from the rest of us. Considering how unpleasant it is to follow a low-calorie diet, if the molecular pathway downstream of CR can be identified, perhaps 'CR mimetics' can be developed which will produce the anti-ageing benefits but spare us the hunger.

The precise mechanism by which CR delays ageing has not been established, but it has been assumed that some global metabolic switch is involved, together with reduced free-radical production. Calorie restriction also reduces blood glucose and insulin levels, suggesting that neuroendocrine signalling and reduced glycation of macromolecules may also be involved. Finally, CR may act as a minor stress, which 'trains' the animal to deal with free-radical and other damage to macromolecules to delay ageing⁵.

In yeast, lifespan is defined by the number of divisions a mother cell can undergo to

generate daughter cells (Fig. 1) before petering out. Lin and colleagues⁶ previously showed that yeast cells cultured in 0.5% glucose produce 30% more generations than cells cultured in 2% glucose, and that this lifespan extension depends on a gene, *SIR2*. This gene encodes an enzyme that removes an acetyl group from (deacetylates) histones, the proteins in which DNA is packaged in the eukaryotic chromosome. It probably exerts its effects on lifespan by suppressing gene transcription and recombination of certain chromosomal regions. The enzyme concerned is regulated by NAD, a form of energy currency in a cell. NAD is a cofactor that transfers high-reducing-potential electrons derived from sugars and other metabolic sources to the electron-transport chain of eukaryotic mitochondria to oxygen to generate ATP, also an energy currency in the cell. NAD levels are high when cells are starved. Sir2 protein activity is high when NAD levels are high, consistent with the idea that Sir2 regulation of gene expression is coupled to metabolic status.

In what appears paradoxical at first, Lin *et al.*³ found that CR yeast respire at a threefold higher rate, as measured by oxygen consumption. Yet this result is not surprising, because yeast can produce energy and grow in two ways, by fermentation or respiration. Normally, when glucose is plentiful, the genes controlling respiration are suppressed, and yeast cells ferment glucose to ethanol, an anaerobic process — one not requiring oxygen. But when glucose is limited, the cells switch to respiration, which uses oxygen and is a far more efficient way of producing energy. Previous work⁷ has shown that there are indeed major shifts in the abundance of messenger RNAs that encode metabolic enzymes when yeast grow in low glucose, indicating a change from fermentation to respiration.

Lin *et al.*³ also demonstrate that respiration is key to the increase in longevity by showing that, under CR conditions, yeast that cannot respire because of a mutation in their electron-transport chain do not live longer than normal. This result suggests that lifespan is increased through the increase in NAD that results from efficient electron transport during respiration, with the NAD increasing Sir2 activity and therefore lifespan. There may be more convoluted signalling pathways involved, but this one is the simplest.

To try to induce the CR gene-expression cascade without the low-glucose diet, the authors produced a genetically engineered yeast strain possessing high levels of the gene *HAP4*, which encodes a transcription factor that activates many respiratory genes⁸. This strain has an increased respiration rate, and lives longer even when grown in high-glucose media; as in CR, the lifespan extension depends on *SIR2*. Lin *et al.* found that some of the changes in gene expression



Figure 1 Yeast in bud — how a mother cell generates daughters.

in the CR yeast — and not just the expression of genes involved in respiration — are similar to those observed with *HAP4* overexpression. These changes may be critical for lifespan extension, but they do not yet point to an obvious key pathway.

Because free radicals are by-products of respiration, cells with a higher respiration rate would be expected to produce more of them. Consistent with this, the authors show that the CR yeast are more sensitive to two free-radical-inducing chemicals, paraquat and hydrogen peroxide; this is generally interpreted as an indication of greater free-radical production by cells. So these findings contradict the theory that a decrease in free radicals is an essential feature of increased longevity⁹. However, complex pathways exist to detoxify free radicals or to reverse damage induced in macromolecules by free radicals (DNA-repair pathways, for instance)¹⁰. In a survey of gene-expression changes of enzymes involved in such functions, including catalase and superoxide dismutase, the authors detected no evidence that these were increased in CR cells. But there are likely to be many other enzymes that protect against free radicals. About 13% of the yeast nuclear genome encodes mitochondrial proteins¹¹, and some of these may be protective.

SIR2 is linked to metabolism via the regulation of NAD levels by respiration and, in turn, by NAD regulation of Sir2-induced histone deacetylation. But there is as yet no connection between Sir2 and free-radical production. Increased yeast lifespan caused by high levels of Sir2 could be occurring without any change in metabolic rate or free-radical output. Perhaps some of the genes regulated by Sir2 encode mitochondrial proteins with protective effects against free radicals, so that they confer long lifespan in the face of increased free-radical production.

What applies in yeast won't necessarily

apply in animals. Notably, animals do not have the luxury of choosing between respiration and anaerobic growth. Under CR, mammals shift metabolism towards the breakdown of fat or glycogen and the biosynthesis of glucose, rather than glucose respiration. It is not clear whether the respiration rate changes in CR mammals, but free-radical-induced cellular damage seems to be reduced⁹. Interestingly, some of the gene-expression changes detected in CR mice, including changes in genes that regulate energy metabolism and stress, are similar to those observed during the shift from anaerobic to aerobic growth in yeast¹². And *SIR2* is known to affect lifespan in the worm *Caenorhabditis elegans* as well as in yeast¹³. There may therefore be a common gene-regulatory cascade downstream of *SIR2* that is involved in regulating lifespan, even though the upstream regulatory pathway differs. Further analysis of the *SIR2* downstream pathway in yeast may well illuminate what happens in mammals. ■

Siu Sylvia Lee and Gary Ruvkun are in the Department of Genetics, Harvard Medical School, and the Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

e-mail: ruvkun@molbio.mgh.harvard.edu

- McCay, C., Crowell, M. & Maynard, L. *J. Nutr.* **10**, 63–79 (1935).
- Finch, C. E. (ed.) in *Longevity, Senescence, and the Genome* 504–536 (Univ. Chicago Press, 1990).
- Lin, S.-J. *et al. Nature* **418**, 344–348 (2002).
- Roth, G. S., Ingram, D. K. & Lane, M. A. *J. Am. Geriatr. Soc.* **47**, 896–903 (1999).
- Masoro, E. *J. Exp. Gerontol.* **35**, 299–305 (2000).
- Lin, S.-J., Defossez, P. A. & Guarente, L. *Science* **289**, 2126–2128 (2000).
- DeRisi, J. L., Iyer, V. R. & Brown, P. O. *Science* **278**, 680–686 (1997).
- Blom, J., De Mattos, M. J. & Grivell, L. A. *Appl. Environ. Microbiol.* **66**, 1970–1973 (2000).
- Sohal, R. S. & Weindruch, R. *Science* **273**, 59–63 (1996).
- Finkel, T. & Holbrook, N. J. *Nature* **408**, 239–247 (2000).
- Kumar, A. *et al. Genes Dev.* **16**, 707–719 (2002).
- Weindruch, R., Kaye, T., Lee, C. K. & Prolla, T. A. *J. Nutr.* **131**, 918S–923S (2001).
- Tissenbaum, H. A. & Guarente, L. *Nature* **410**, 227–230 (2001).

Cacao usage by the earliest Maya civilization

Foaming chocolate prepared in spouted vessels made a delectable Preclassic drink.

The Maya archaeological site at Colha in northern Belize, Central America, has yielded several spouted ceramic vessels that contain residues from the preparation of food and beverages. Here we analyse dry residue samples by using high-performance liquid chromatography coupled to atmospheric-pressure chemical-ionization mass spectrometry, and show that chocolate (*Theobroma cacao*) was consumed by the Preclassic Maya as early as 600 BC, pushing back the earliest chemical evidence of cacao use by some 1,000 years. Our application of this new and highly sensitive analytical technique could be extended to the identification of other ancient foods and beverages.

The site at Colha is known for its specialized production of lithic tools¹ and for its collection of intact, spouted vessels², which were manufactured only during the Preclassic period (900 BC to AD 250)³. These have generally been recovered from burial sites associated with elite individuals and are relatively rare, typically being found with other serving vessels such as bowls, dishes and plates, and were probably used to dispense liquid from the spout, in much the same way as a teapot.

Based on epigraphic analysis of vessels dating from the Classic (AD 250–900) period and documents written at the time of the Spanish Conquest, liquid chocolate was frothed to produce a foam — considered by the Maya and the Aztecs to be the most desirable part of the drink^{4,5} — by pouring the liquid from one vessel into another⁴. In the earlier Preclassic spouted jars (Fig. 1), frothing would also have been accomplished by introducing air through the spout,

enabling the vessel to be used for preparing, as well as pouring, liquid chocolate.

At the time of the Spanish Conquest, chocolate was consumed with most meals and was usually mixed with another ingredient (for example, water, maize, chilli and/or honey) and in different proportions to produce a variety of drinks⁵. Our aim, however, was to confirm the existence of cacao residues in spouted vessels, rather than to investigate the presence of other components.

We used high-performance liquid chromatography (HPLC) coupled to atmospheric-pressure chemical-ionization mass spectrometry (APCI MS) to analyse each of the samples collected from the 14 vessels recovered from a series of burials at Colha^{6,7}. All of these vessels date to between 600 BC and AD 250 (ref. 2). The procedure involved withdrawing about 500 mg from each sample vial, and adding 3 ml distilled water at 80 °C to solubilize the materials. Before analysis, we passed each sample through membrane filters to eliminate particulate matter.

Cacao has a unique chemical composition of over 500 different compounds, including members of the methylxanthine class (primarily theobromine, with a lower concentration of caffeine). As *T. cacao* is the only Mesoamerican plant that contains theobromine as the primary methylxanthine⁸, this compound can be used as a marker for the presence of cacao. For example, HPLC coupled to thermospray mass spectroscopy has revealed cacao residues in ceramic vessels found in an Early Classic (AD 460–480) tomb at the Maya site of Rio Azul in northeastern Guatemala^{8–10}.



Figure 1 Early Maya use of cacao (*Theobroma cacao*): spouted vessel no. 13, which was found to contain cocoa residue. This vessel is one of 14 excavated from Colha in northern Belize, dating to between 600 BC and AD 250.

For APCI MS, the probe was operated in positive-ion mode to monitor for peaks at $m/z = 181$ (theobromine) and $m/z = 195$ (caffeine), with the ultraviolet detector set at 270 nm. The results of the HPLC MS confirmed the existence of theobromine in 3 of the 14 samples analysed. Peaks from the extract of the residue from vessel 13 (Fig. 1) were evident in the total-ion and ultraviolet chromatograms and the selected-ion-monitoring trace at $m/z = 181$ (Fig. 2; for peak calibration for standards, see supplementary information) and was confirmed by the fact that the mass spectrum (see supplementary information) and ultraviolet chromatogram of this peak show the same retention time as theobromine (Fig. 2b).

To our knowledge, this is the first time that this new technique has been used to analyse dry residues from the interior surfaces of prehistoric pottery. The presence of cacao in Maya spouted vessels at Colha indicates that its usage pre-dates evidence from Rio Azul by almost a millennium. We now know that the Maya had a long, continuous history of preparing and consuming liquid chocolate from the Preclassic period through to the Spanish Conquest. Cacao wood charcoal dating to the same period has been found at several sites in the region¹¹, further supporting the idea that cacao drinking has its roots in the Preclassic, and indicating that this part of northern Belize may have been one of the main production areas for cacao during this period.

W. Jeffrey Hurst*, Stanley M. Tarka Jr*, Terry G. Powis†, Fred Valdez Jr†, Thomas R. Hester†

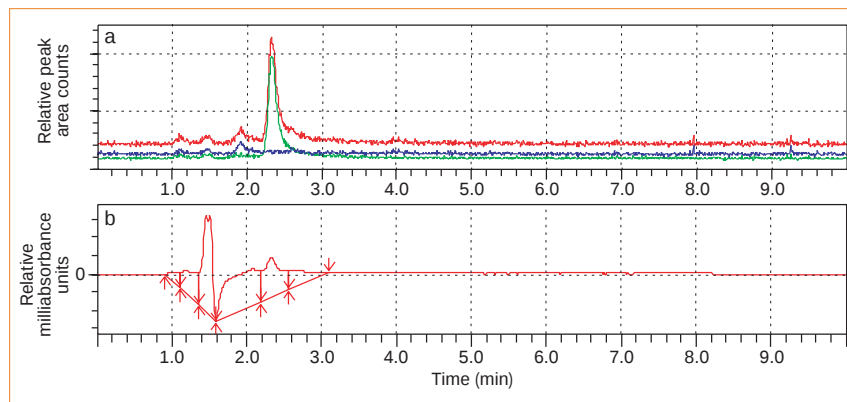


Figure 2 Chromatographic analysis of residue from Maya cooking vessels. **a**, Mass-spectrometry chromatogram of dry extract from vessel 13. Red, total-ion chromatogram; green, selected-ion monitoring (SIM) at $m/z = 181$ for theobromine; blue, SIM at $m/z = 195$ for caffeine. For the mass spectrum of the peak at 2.6 min, as well as details of methods and standards, see supplementary information. **b**, Ultraviolet chromatogram of the same extract, showing a peak at 2.6 min, which is consistent with results in **a**. Dashed lines are integration markers; arrows indicate the starts and ends of peaks. The peak at around 1.5 min is due to the chromatography solvent. The peak for caffeine is not evident as its concentration in cocoa is about 10 times lower than that of theobromine.

*Hershey Foods Technical Center, PO Box 805,

Hershey, Pennsylvania 17033, USA

e-mail: whurst@hersheys.com

†Department of Anthropology, University of Texas,

Austin, Texas 78712-1086, USA

1. Hester, T. R. & Shafer, H. J. in *Archaeological Views from the Countryside: Village Communities in Early Complex Societies* (eds Schwartz, G. M. & Falconer, S. E.) 48–63 (Smithsonian Institution, Washington DC, 1994).
2. Valdez, F. Jr *The Prehistoric Ceramics of Colha, Northern Belize*. Thesis, Harvard Univ. (1987).
3. Powis, T. G. & Hurst, W. J. *Proc. 66th Annu. Meeting Soc. Am. Archaeol.* (New Orleans, 2001).
4. Coe, S. D. & Coe, M. D. *The True History of Chocolate* (Thames & Hudson, London, 1996).
5. Tozzer, A. M. *Landa's Relación de Las Cosas de Yucatán* (Kraus

Reprint, New York, 1941).

6. Potter, D. R. in *The Colha Project, Second Season, 1980 Interim Report* (eds Hester, T. R., Eaton, J. D. & Shafer, H. J.) 173–184 (Center for Archaeological Research, San Antonio, Texas; Centro Studi Ricerche Ligabue, Venice; 1980).
7. Potter, D. R. in *Archaeology at Colha, Belize, 1981 Interim Report* (eds Hester, T. R., Shafer, H. J. & Eaton, J. D.) 98–122 (Center for Archaeological Research, San Antonio, Texas; Centro Studi Ricerche Ligabue, Venice; 1982).
8. Hurst, W. J., Martin, A. J. Jr, Tarka, S. M. Jr & Hall, G. D. *J. Chromatogr.* **466**, 279–289 (1989).
9. Hall, G. D., Tarka, S. M. Jr, Hurst, W. J., Stuart, D. & Adams, R. E. W. *Am. Antiquity* **55**, 138–143 (1990).
10. Stuart, D. *Antiquity* **62**, 153–157 (1988).
11. Turner, B. L. & Miksicsek, C. H. *Econ. Bot.* **38**, 179–193 (1984).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Cell biology

Targeted transfection by femtosecond laser

The challenge for successful delivery of foreign DNA into cells *in vitro*, a key technique in cell and molecular biology with important biomedical implications, is to improve transfection efficiency while leaving the cell's architecture intact. Here we show that a variety of mammalian cells can be directly transfected with DNA without perturbing their structure by first creating a tiny, localized perforation in the membrane using ultrashort (femtosecond), high-intensity, near-infrared laser pulses. Not only does this superior optical technique give high transfection efficiency and cell survival, but it also allows simultaneous evaluation of the integration and expression of the introduced gene.

Previous techniques that have been developed for transfection of cells with DNA¹ include carrier-mediated transfer² and transfer by plasma-membrane permeabilization³, as well as direct transfer⁴, but the efficiency of targeted DNA delivery by these methods may not be optimal. Moreover, none allows contact-free, non-disruptive, stable transfection of individual cells and concomitant evaluation *in situ* of transgene expression.

We directed a high-intensity (10^{12} W cm⁻²), near-infrared, femtosecond-pulsed laser beam (wavelength, 800 nm) from an 80-MHz titanium-sapphire laser, with a mean power of 50–100 mW and tightly focused using a high-numerical-aperture objective, at a sub-femtolitre focal volume at the cell membrane. This resulted in the formation of a single, site-specific, transient perforation in the cell membrane through which DNA could enter. This mode of targeted transfection differs from the less precise nanosecond-pulsed, ultraviolet (355 nm) lasers used previously⁵ and which were found to disrupt cellular integrity⁶.

Using Chinese hamster ovarian (CHO) and rat-kangaroo kidney epithelial (PtK2) cells, we studied the process of transfection

mediated by intense near-infrared femtosecond laser pulses. Cells were suspended inside a sterile miniaturized cell chamber in 0.5 ml culture medium containing 0.2 µg plasmid DNA vector pEGFP-N1 (4.7 kilobases) encoding enhanced green fluorescent protein (EGFP)⁷. Transmission images of cells were obtained at low power (< 5 µW), and the near-infrared laser beam was then focused (under the same microscope) on the edge of the membrane of a target cell, which was exposed to an enhanced mean laser power of 50–100 mW for 16 ms so that transfection could occur. More than 200 cells of each type were targeted in each of 18 replicate experiments; it took 10–15 s to prepare for the transfection of each cell.

We assessed the integration and expression efficiency of the EGFP gene *in situ* by time-lapse two-photon fluorescence imaging⁸ at a mean laser power of < 1 mW over a period of 72 h, as well as by two-photon fluorescence-lifetime imaging (TPFLIM)⁹. Figure 1 shows that diffraction-limited focusing of intense femtosecond near-infrared laser pulses selectively facilitates transfection of the target cells, but not of the adjacent cells. Expression of EGFP in the transfected cell is also demonstrated by TPFLIM, and the measured fluorescence lifetime of about 2.4 ns is consistent with that reported for mammalian cells expressing EGFP¹⁰.

Irrespective of cell type, the transfection achieved by this technique was invariably 100%. This high level of selective and total transfection, without any detrimental effects on growth and division, and with virtually no cell death or sign of apoptosis, together with the ability to determine expression by fluorescence-intensity imaging and TPFLIM with the same microscope, demonstrate the potential of this non-disruptive technique in transfection and expression studies. The ability to transfer foreign DNA safely and efficiently into specific cell types (including stem cells) — circumventing the need for mechanical, electrical or chemical means — will be an encouraging advance for a range of ventures, including targeted gene therapy and DNA vaccination.

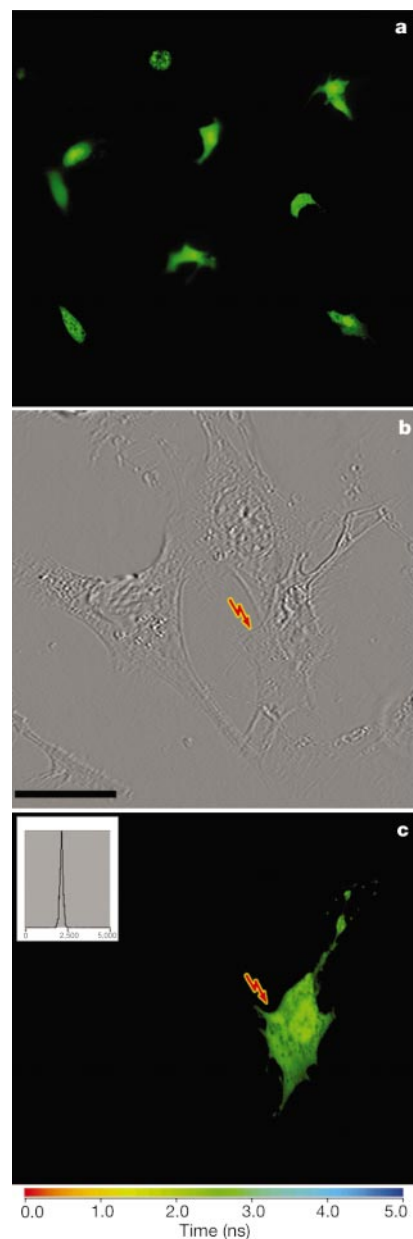


Figure 1 Analysis of the targeted transfection of Chinese hamster ovarian (CHO) cells with a plasmid encoding enhanced green fluorescent protein (EGFP) by *in situ* visualization, and measurement of its expression by near-infrared, two-photon-excitation-evoked, real EGFP fluorescence detection and fluorescence-lifetime imaging. **a**, Real EGFP fluorescence image of several CHO cells transfected with the pEGFP-N1 plasmid. **b**, Transmission image of a single transfected CHO cell (arrow). **c**, Two-photon fluorescence-lifetime image of the same cell expressing EGFP; colour scale indicates the fluorescence lifetime (τ) between 0 and 5 nanoseconds. Inset, distribution of fluorescence lifetime (in picoseconds) of EGFP throughout the entire transfected cell. Scale bar, 25 µm.

Uday K. Tirlapur, Karsten König

Laser Microscopy Division, Institute of Anatomy II, Friedrich Schiller University, Teichgraben 7, 07743 Jena, Germany

e-mail: utir@mti-n.uni-jena.de

1. Stephens, D. J. & Pepperkok, R. *Proc. Natl Acad. Sci. USA* **98**, 4295–4298 (2001).

2. Luo, D. & Saltzman, M. T. *Nature Biotechnol.* **18**, 33–37 (2000).
3. Bildirici, L. *et al.* *Nature* **405**, 298 (2000).
4. Knoblauch, M. *et al.* *Nature Biotechnol.* **17**, 906–909 (1999).
5. Tao, W. *et al.* *Proc. Natl Acad. Sci. USA* **84**, 4180–4184 (1987).
6. König, K. J. *Microsc.* **200**, 83–104 (2000).
7. Lippincott-Schwartz, J. *et al.* *Nature Rev. Mol. Cell Biol.* **2**, 444–456 (2001).
8. Squirrell, J. M. *et al.* *Nature Biotechnol.* **17**, 763–767 (1999).
9. So, P. T. C. *et al.* *Biomaging* **3**, 1–15 (1995).
10. Pepperkok, R. *et al.* *Curr. Biol.* **9**, 269–272 (1999).

Competing financial interests: declared none.

Neurodegenerative disease

Amyloid pores from pathogenic mutations

Alzheimer's and Parkinson's diseases are associated with the formation in the brain of amyloid fibrils from β -amyloid and α -synuclein proteins, respectively. It is likely that oligomeric fibrillization intermediates (protofibrils), rather than the fibrils themselves, are pathogenic, but the mechanism by which they cause neuronal death remains a mystery. We show here that mutant amyloid proteins associated with familial Alzheimer's and Parkinson's diseases form morphologically indistinguishable annular protofibrils that resemble a class of pore-forming bacterial toxins, suggesting that inappropriate membrane permeabilization might be the cause of cell dysfunction and even cell death in amyloid diseases.

The possibility that a molecular species other than the amyloid fibril could be pathogenic arose when oligomeric species rich in β -sheet structure (protofibrils) were found to be discrete intermediates in the fibrillization of β -amyloid (A β) and of α -synuclein *in vitro*^{1,2}. An intermediate protofibril might therefore be pathogenic and be 'detoxified' by conversion to a fibril, as suggested by three general observations: there is no correlation between the quantity of fibrillar deposits at autopsy and the clinical severity of Alzheimer's or Parkinson's disease; transgenic mouse models of these conditions have disease-like phenotypes before fibrillar deposits can be detected^{1,2}; and non-fibrillar A β oligomers are toxic in cell culture^{3,4} and have activity *in vivo*⁵.

Toxic protofibrils have been implicated in other neurodegenerative diseases as well as in systemic amyloidoses such as type II diabetes (in which the amyloid protein is IAPP)⁶ and familial amyloidotic polyneuropathy (in which it is transthyretin)⁷. Strikingly, protofibrils comprising proteins that are not associated with any disease are also toxic, suggesting that toxicity might arise from a shared structural feature of these intermediates⁸.

The pathogenic-protofibril hypothesis is supported by biophysical studies of variants of A β and α -synuclein linked to autosomal-dominant forms of Alzheimer's and Parkinson's

diseases, respectively. The 'Arctic' mutation in amyloid-precursor protein, unlike all other mutations associated with Alzheimer's disease, reduces the total concentration of circulating A β , but mutant A β_{ARC} forms protofibrils *in vitro* more rapidly and to a greater extent than the wild-type form⁹.

The A30P and A53T α -synuclein mutations associated with Parkinson's disease (in which an alanine residue is replaced by phenylalanine at position 30 or by threonine at position 53, respectively) both promote protofibril formation *in vitro* relative to wild-type α -synuclein². We examined the structural properties of A30P, A53T and A β_{ARC} protofibrils for shared structural features that might be related to their toxicity.

Heterogeneous populations of A30P, A53T and A β_{ARC} protofibrils were fractionated by gel-filtration chromatography (H.A.L. *et al.*, unpublished results). The fraction with the smallest A30P and A53T protofibrils contained β -sheet-rich (as measured by circular dichroism) oligomers comprising 20–25 α -synuclein molecules (relative molecular mass, 320K–380K; 22–26 monomers; H.A.L. *et al.*, unpublished results). Analysis of this fraction by electron microscopy revealed annular species (diameter 8–12 nm; inner diameter 2.0–2.5 nm) and coiled species, both of which seemed to be related to the spherical and chain-like species reported earlier². Protofibrillar A β_{ARC} , fractionated by a similar method, contained many annular species of similar appearance (Fig. 1), diameter (7–10 nm; inner diameter 1.5–2.0 nm) and relative molecular mass (150K–250K, 40–60 A β_{ARC} molecules).

The pore-like morphology of a subpopulation of amyloid protofibrils might explain the pore activity of α -synuclein protofibrils in vesicle-permeabilization models¹⁰ and the

channel-like properties of A β (ref. 11). Other amyloid proteins, including huntingtin (in Huntington's disease)¹² and IAPP^{6,12}, also have pore-like activity *in vitro*. The small annular A β and α -synuclein protofibrils (Fig. 1) resemble the cytolytic β -barrel pore-forming toxins from bacteria such as *Clostridium perfringens*¹³.

As expected, amyloid pores are formed much less efficiently than bacterial pores, which during the course of evolution have optimized their ability to puncture host membranes. However, amyloid pores might be wholly or partly responsible for the cytotoxicity associated with the formation of amyloid fibrils in Alzheimer's and Parkinson's diseases and in other age-associated degenerative amyloid diseases.

Hilal A. Lashuel*, Dean Hartley*, Benjamin M. Petre†, Thomas Walz†, Peter T. Lansbury Jr*

**Center for Neurologic Diseases, Brigham and Women's Hospital and Department of Neurology, Harvard Medical School, 65 Landsdowne Street, Cambridge, Massachusetts 02139, USA*
e-mail: plansbury@rics.bwh.harvard.edu

†Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA

1. Lansbury, P. T. *Proc. Natl Acad. Sci. USA* **96**, 3342–3344 (1999).
2. Goldberg, M. S. & Lansbury, P. T. *Nature Cell Biol.* **2**, E115–E119 (2000).
3. Hartley, D. M. *et al.* *J. Neurosci.* **19**, 8876–8884 (1999).
4. Lambert, M. P. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 6448–6453 (1998).
5. Walsh, D. M. *et al.* *Nature* **416**, 535–539 (2002).
6. Janson, J. *et al.* *Diabetes* **48**, 491–498 (1999).
7. Sousa, M. M. *et al.* *Am. J. Pathol.* **159**, 1993–2000 (2001).
8. Bucciantini, M. *et al.* *Nature* **416**, 507–511 (2002).
9. Nilsberth, C. *et al.* *Nature Neurosci.* **4**, 887–893 (2001).
10. Volles, M. J. & Lansbury, P. T. *Biochemistry* **41**, 4595–4602 (2002).
11. Lin, H., Bhatia, R. & Lal, R. *FASEB J.* **15**, 2433–2444 (2001).
12. Kagan, B. L., Hirakura, Y., Azimov, R. & Azimova, R. *Brain Res. Bull.* **56**, 281–284 (2001).
13. Hotze, E. M. *et al.* *J. Biol. Chem.* **277**, 11597–11605 (2002).

Competing financial interests: declared none.

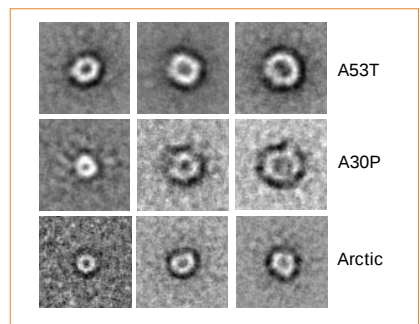


Figure 1 Mutations that cause Alzheimer's and Parkinson's diseases promote the formation of amyloid pores. Projection averages of annular protofibrils formed by the Parkinson's-disease-linked α -synuclein mutants A53T and A30P and by the Alzheimer's-disease-linked A β (1–40)_{ARC} 'Arctic' mutant (E22G). The images were calculated from a total of about 5,000–6,000 particles, which were obtained from 25–31 digitized electron micrographs of purified (by gel filtration on Superdex-200 (α -synuclein) or Superose-6 (A β)) protofibrillar fractions corresponding to the lowest-*M_r* fraction of each mutant. Each panel shows an original area of 30.5 × 30.5 nm.

COMMUNICATIONS ARISING

Climate change

Recent temperature trends in the Antarctic

It is important to understand how temperatures across the Antarctic have changed in recent decades because of the huge amount of fresh water locked into the ice sheet and the impact that temperature changes may have on the ice volume. Doran *et al.*¹ claim that there has been a net cooling of the entire continent between 1966 and 2000, particularly during summer and autumn. We argue that this result has arisen because of an inappropriate extrapolation of station data across large, data-sparse areas of the Antarctic.

The Antarctic-wide analysis made use of the University of East Anglia HadCRUT

temperature data set, which provides monthly temperature anomalies (with respect to 1961–90) for $5^\circ \times 5^\circ$ grid boxes based on land and ship reports². A spatially weighted average of the HadCRUT grid boxes that have data from 1958 to 2000 gives a warming of 0.176°C per decade, which is significant at the 95% level.

Doran *et al.* calculated spatially smoothed trends using a technique that allowed grid boxes (there are only 16 of these south of 65°S) to have a radius of influence of roughly 25% of the maximum width of the image, which we believe is too large, considering the highly localized nature of the factors that influence the climate at many of the stations. We suggest that the interpolation has given too much weight to grid boxes in data-sparse regions, resulting in a misleading representation of cooling over the continent, which is not supported by *in situ* or remote-sensing data.

Doran and colleagues' Fig. 2 shows annual and seasonal temperature trends for 1966–2000, with the largest cooling occurring in autumn over an area from the southern Weddell Sea to the South Pole. There is also pronounced cooling in spring and in annual trends. However, their Fig. 2 does not show the marked warming on the western side of the Antarctic Peninsula, which is greatest during winter³. It is unclear why the authors chose 1966–2000 for their analysis, as most of the temperature series in the HadCRUT data set begin in the late 1950s.

Although the spacing of stations around the coast is reasonable in the eastern hemisphere, there are large gaps in the data for the coast of west Antarctica, and only two stations in the interior have long records: those at the South Pole and Vostok (78.5°S , 106.9°E). The area in which Doran *et al.* report the greatest cooling is devoid of stations with long records, and HadCRUT includes no data for this region.

The warming on the western side of the Antarctic Peninsula is of limited spatial extent and is greatest close to Faraday Station (65.3°S , 64.3°W), where the trend for 1951–2000 is an increase of 1.09°C per decade during winter and 0.56°C per decade annually (both values are significant at the 95% level). However, at Halley Station (75.5°S , 26.4°W), where there is a continuous record dating back to 1957, there is an indication of a slight cooling over this period during autumn, but a small warming during other seasons, although none of these trends is statistically significant. The South Pole shows limited cooling in each season, although only the annual trend of 0.20°C of cooling per decade is statistically significant at the 90% level. Data from Vostok for 1957–2000 do not show any statistically significant trend in any season or in the annual data.

Trends in the ice-skin temperature based on remote-sensing data have been examined⁴ for 1979–98. The greatest cooling was found over the high plateau of east Antarctica, with some cooling over west Antarctica, but with warming over the area from the southern Weddell Sea towards the South Pole. Another study⁵ examined temperature trends across west Antarctica using automatic weather-station observations and satellite passive microwave measurements: the only statistically significant trend found at the 95% level was an increase of 2°C at Siple (75.9°S , 84.2°W) over the period 1979–97.

Attempting to derive a temperature trend for the entire Antarctic continent is almost meaningless, as huge areas are devoid of long-term, *in situ* climate measurements. At present, the trends from the various stations present a spatially complex picture of change across the continent during recent decades and do not indicate any consistent warming or cooling.

**John Turner*, John C. King*,
Tom A. Lachlan-Cope*, Phil D. Jones†**

**British Antarctic Survey, Natural Environment Research Council, Cambridge CB3 0ET, UK
e-mail: j.turner@bas.ac.uk*

†Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, UK

1. Doran, P. T. *et al.* *Nature* **415**, 517–520 (2002).
2. Jones, P. D., New, M., Parker, D. E., Martin, S. & Rigor, I. G. *Rev. Geophys.* **37**, 173–199 (1999).
3. Vaughan, D. G., Marshall, G. J., Connolley, W. M., King, J. C. & Mulvaney, R. *Science* **293**, 1777–1779 (2001).
4. Comiso, J. C. *J. Clim.* **13**, 1674–1696 (2000).
5. Shuman, C. A. & Stearns, C. R. *J. Clim.* **14**, 1977–1988 (2001).

Doran et al. reply — Turner *et al.* do not find fault with our main focus — the rapid ecological response to recent cooling in the McMurdo Dry Valleys. The essence of their comment is that the spatial interpolation of the Antarctic continental data set (our Fig. 2) does not provide a meaningful picture of recent temperature trends. Although any interpolation is open to question, we note the following points.

First, weighting by inverse-fourth-power of separation distance in our analysis effectively eliminates contributions near to the 2,284-km radius of influence if other stations are closer. Only between the Antarctic Peninsula and the Ross Sea does the gap between points approach the radius of influence. Second, the trend maps in our Fig. 2 have features that are finer than the quarter-image radius of influence, indicating that that local data prevail in our analysis. Third, the interpolated fields are consistent with the summaries for 1976–2000 in Fig. 2.10 of the recent IPCC WG-I report¹ and with trends based on a different data set² for 1965–1999, an interval that was chosen because it was

one of global warming. A recent depiction³ of combined summer and autumn Antarctic surface-temperature trends from 1969 to 2000 is similar to ours, although our data suggest that cooling is more pronounced during autumn periods than in summer.

The key question that arises is this: is our interpolation better than arithmetic averaging? We contend that, although the interpolations involve uncertainty, they highlight the fact that a full assessment of Antarctic temperature trends requires more than the averaging schemes that have so far been used to imply that Antarctica has been warming at a rate that is faster than the global average^{4,5}.

In estimates of hemispheric anomalies or trends (for example, see the IPCC's Figure 2.7; ref. 1), regions that are devoid of data are effectively assigned anomalies (or trends) that are equal to the hemispheric means of anomalies (or trends) for areas for which data exist. This type of assignment ignores information from even nearest-neighbour grid cells. We maintain that our approach represents an improvement over arithmetic means of station values or grid cells, provided that the nature of the interpolation procedure is clearly stated.

Only by interpolation can one hope to determine the area fraction and spatial pattern of continental warming or cooling, regardless of the magnitudes of such trends — which are irrelevant to our conclusions. As Antarctic trends obviously vary spatially, seasonally and interdecadally, interpolation will ultimately be required to optimize the information contained in historical data.

**J. E. Walsh, P. T. Doran*, J. C. Prisco,
W. B. Lyons, A. G. Fountain,
D. M. McKnight, D. L. Moorhead,
R. A. Virginia, D. H. Wall, G. D. Clow,
C. H. Fritsen, C. P. McKay, A. N. Parsons**

**Department of Earth and Environmental Sciences, University of Illinois at Chicago, Illinois 60607, USA
e-mail: pdoran@uic.edu*

1. Houghton, J. T. *et al.* (eds) *Climate Change 2001: The Scientific Basis* (Intergovernmental Panel on Climate Change, Cambridge Univ. Press, 2001).
2. Hansen, J. *et al.* *J. Geophys. Res.* **106** (D20), 23947–23963 (2001).
3. Thompson, D. J. W. & Solomon, S. *Science* **296**, 895–899 (2002).
4. Jacka, T. H. & Budd, W. F. *Ann. Glaciol.* **27**, 553–559 (1998).
5. Vaughan, D. G., Marshall, G. J., Connolley, W. M., King, J. C. & Mulvaney, R. *Science* **293**, 1777–1779 (2001).

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com/nature) or on request from nature@nature.com

A molecular programme for the specification of germ cell fate in mice

Mitinori Saitou, Sheila C. Barton & M. Azim Surani

Wellcome Trust/Cancer Research UK Institute of Cancer and Developmental Biology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QR, UK

Germ cell fate in mice is induced in proximal epiblast cells by the extra-embryonic ectoderm, and is not acquired through the inheritance of any preformed germ plasm. To determine precisely how germ cells are specified, we performed a genetic screen between single nascent germ cells and their somatic neighbours that share common ancestry. Here we show that *fragilis*, an interferon-inducible transmembrane protein, marks the onset of germ cell competence, and we propose that through homotypic association, it demarcates germ cells from somatic neighbours. Using single-cell gene expression profiles, we also show that only those cells with the highest expression of *fragilis* subsequently express *stella*, a gene that we detected exclusively in lineage-restricted germ cells. The *stella* positive nascent germ cells exhibit repression of homeobox genes, which may explain their escape from a somatic cell fate and the retention of pluripotency.

August Weismann proposed that embryonic cells that inherit preformed germ plasm acquire germ cell fate while the rest become somatic cells¹, which is the case in some organisms^{2–4}. In mice, however, there is an apparent lack of such preformed germ cell determinants⁵. Instead, germ cell competence is induced in the proximal epiblast cells of the egg cylinder at embryonic day (E)6.5 in response to signals, including Bmp4 and Bmp8b, from the extra-embryonic ectoderm^{6–12}. However, these proximal epiblast cells are not lineage restricted, as they give rise to primordial germ cells (PGCs) as well as somatic cells, including the extra-embryonic mesoderm and allantois⁵. How then precisely are PGCs specified from precursor cells that can at the same time give rise to somatic cells, and what are the intrinsic differences between the nascent PGCs, somatic cells and the original epiblast cells? A mechanism must exist to prevent the founder population of about 45 PGCs at E7.2 from acquiring a somatic cell fate.

The genetic basis for the specification of germ cell fate in mice is not well defined. At present, PGCs are distinguished primarily by their characteristic tissue nonspecific alkaline phosphatase (TNAP) staining^{13,14}, but TNAP is also present in somatic cells that surround them¹⁵. Even *Oct4*, a gene that is primarily expressed in pluripotent lineages and regarded to have a role in germ cell development^{16–18}, shows widespread expression in early embryos.

A particular difficulty in addressing this question has been that there are only a few founder PGCs, and these are embedded among somatic cells. We therefore decided to approach this problem through single cell analysis of founder PGCs and their somatic neighbours. This analysis has provided a detailed insight into the molecular events for the specification of germ cell fate. At the centre of these events are two genes, *fragilis* and *stella*, which we propose are involved in initiating germ cell competence and specification, and in the demarcation of PGCs from their somatic neighbours.

Single-cell complementary DNAs from nascent PGCs

First, we carried out a careful analysis of late streak/early bud stage (E7.0–E7.5) embryos¹⁹ from the 129/SvEv mouse strain for the location of the TNAP positive founder PGCs (Fig. 1a, b). Then we dissected an early bud stage embryonic fragment made up of 250–300 cells from the base of the allantois. This fragment consisted of allantoic mesoderm, PGCs and the surrounding mesoderm cells, from which we generated single cells by treatment with trypsin. However, these isolated single cells proved to be morphologically indistinguishable from each other. We therefore picked individual cells randomly and constructed cDNAs from each of them, using a

protocol^{20–22} that we first verified on single embryonic stem cells (ES cells) for representative amplification of cDNAs (data not shown).

We analysed 83 single cell cDNAs, first by using *Bmp4* as a negative marker primarily to identify the amniotic and allantoic mesoderm cells (Fig. 1d). We found 57 out of 83 *Bmp4* positive cells, which we set aside at this stage. Of the remaining 26 cells, 22 were *Tnap* positive (Fig. 1d), and 10 of these cells in particular showed uniformly stronger *Tnap* expression by Southern blotting, which we considered to be of PGC origin (Fig. 1c, d). To confirm this observation, we used expression of the homeobox gene *Hoxb1* (ref. 23), which is thought to be detected only in the mesoderm cells surrounding the PGCs (K. Lawson, personal communication). Indeed, 12 samples were positive for *Hoxb1*, which matched with those with low *Tnap* expression, and therefore were designated as somatic cells (Fig. 1c, d).

Next, we constructed a library from a single PGC cDNA, followed by a differential screen with the PGC and somatic cell cDNA probes. Among 3,000 clones screened initially, 280 (9.3%) were expressed differentially, which were then examined stringently after the inserts were amplified, blotted and probed again with either the PGC or somatic cell cDNA probes. This resulted in 38 (1.3%) candidates, which when screened again with the second PGC and somatic cell probes, resulted in 20 (0.7%) candidates representing 11 different genes by sequence analysis. Further analysis by polymerase chain reaction (PCR) in another two candidate PGC cDNAs and three somatic cell cDNAs revealed two genes with strong or exclusive expression in PGCs that are described below.

fragilis, the first gene that we identified (GenBank accession number AY082484), codes for a protein of 137 amino acids (relative molecular mass of 15,000) with two transmembrane domains, the amino and carboxy terminus ends being located outside the cell. *fragilis* is a newly discovered member of the interferon (IFN)-inducible transmembrane protein family^{24,25}, and is detected in expressed sequence tags (ESTs) derived from many different embryonic and adult tissues, suggesting that it may have a common role in different developmental contexts. One prototype member is the IFN-inducible human 9-27 (identical to the Leu-13 antigen) protein in leukocytes and endothelial cells, a cell surface component of a multimeric complex involved in homotypic adhesion and transduction of antiproliferative signals^{24,26,27}. The second gene, *stella* (GenBank accession number AY082485), codes for a protein of 150 amino acids (relative molecular mass of 18,000), with no sequence similarity to any known protein, and is detected in ESTs from the pre-implantation embryo and germ line, indicating its

predominant expression in totipotent and pluripotent cells. It contains several nuclear localization consensus sequences and a nuclear export signal suggesting that it may shuttle between the nucleus and cytoplasm. *Stella* is highly basic (pI = 9.67; the content of basic residues is 23.3%), suggesting a possible affinity with DNA.

In its N terminus is a modular domain with sequence similarity with the SAP motif²⁸, which is a putative DNA-binding domain involved in chromosomal organization. There is also apparently a splicing factor motif-like structure in its C terminus. These findings suggest a possible involvement of *stella* in chromosomal organization and RNA processing.

Involvement of *fragilis* and *stella* in PGC specification

Next, we sought evidence for the involvement of the two genes in the specification of PGCs. We found strong *fragilis* expression at the base of the incipient allantois, the very region where the founder PGC cluster is observed at E7.25 (Fig. 2a). Only a weak signal was observed in the mesodermal portion of the posterior amnion, and, importantly, the expression did not extend to the allantois. *fragilis* expression persisted until the late bud stage (E7.5) (Fig. 2b), but gradually faded around the early head fold stage (E7.75), and was substantially reduced in migrating PGCs at E8.5 (data not shown).

At an earlier stage, only weak *fragilis* expression was seen throughout the epiblast in E6.0 embryos. But around E6.25–E6.5 (before gastrulation), marked expression was evident within the most proximal layer of the epiblast that is in intimate contact with the extra-embryonic ectoderm, the source of signalling molecules including *Bmp4* and *Bmp8b* (Fig. 2c)^{6,10}. As described below, expression of *fragilis* is indeed induced by extra-embryonic ectoderm through signalling molecules. After gastrulation commenced, the domain of expression migrated to the posterior end of the developing primitive streak at the early/mid streak stage and became very intense in the position where PGCs differentiate from late streak stage onward (Fig. 2d). *fragilis* expression initially extends to a more proximal region, which probably encompasses some of the early allantoic bud cells close to the region. This expression of *fragilis* is however weaker than that observed where the PGC cluster itself resides, and later declines towards its periphery. Notably, this expression pattern is consistent with the origin and migration of PGC precursors as defined by clonal analysis⁵. *fragilis* is therefore the first gene to mark the onset of germ cell competence, and it follows a pattern consistent with the eventual segregation of germ cells from allantoic precursors.

stella, by contrast, was first detected at late streak stage at approximately E7.0 (Fig. 2f). The location, timing and specificity of *stella* signal coincided markedly with the appearance of nascent PGCs. When compared with *fragilis* expression at a very similar developmental stage (Fig. 2e), *stella* expression was seen only in a subset of cells at the centre of *fragilis* positive cells. At the late bud stage (E7.5), the intensity of the *stella* signal increased and some cells considered to be migrating from the initial cluster were also stained (Fig. 2g). At E8.5, groups of migrating PGCs were stained exclusively in the developing hindgut (Fig. 2h). These findings strongly indicate that *stella* is the first gene known to be expressed in the lineage-restricted nascent PGCs, and thereafter in the germ cell lineage.

Next, dissociated cells from the allantoic bud fragment (E7.25) were examined by immunostaining. *Fragilis* was detected in the plasma membrane as well as in the presumptive Golgi apparatus, suggesting rapid synthesis and transport of *fragilis* to the cell surface. Furthermore, *fragilis* was linearly concentrated at cell to cell contact sites (Fig. 3a), consistent with it being an IFN-inducible cell surface protein^{24,26,27}. In contrast, *stella* was seen both in the cytoplasm and the nucleus (Fig. 3b), with the nuclear staining to some degree being complementary to that of propidium iodide, suggesting that *stella* may localize at nuclear matrix structures. We found that the number of *fragilis* positive cells was consistently larger than the *stella* positive cells. Double staining with antibodies revealed that only the cells that apparently exhibit high expression of *fragilis* were positive for *stella* (Fig. 3c). This conclusion was confirmed by the detailed molecular analysis of single nascent PGCs and the somatic neighbours (see below). *Stella* was also

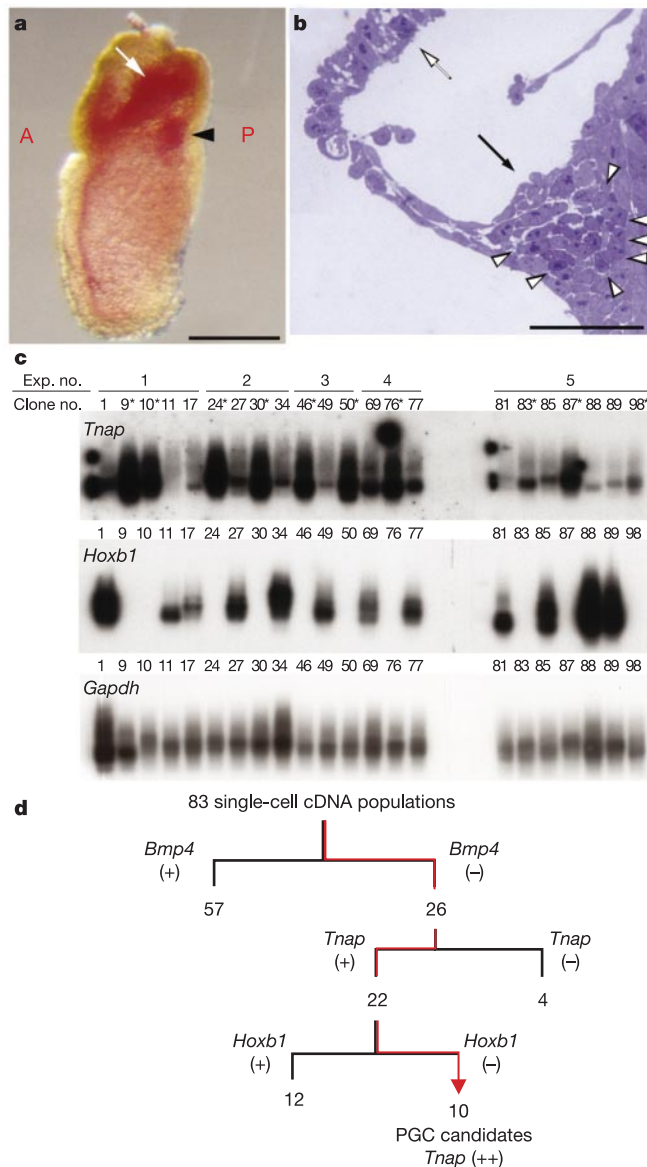


Figure 1 Organization of nascent PGCs. **a**, Late streak (E7.25) 129/SvEv embryo stained for alkaline phosphatase (TNAP), which indicates the region with PGCs (arrowhead). Strong staining is also detected in the chorionic membrane (arrow), and low staining in the epiblast. A, anterior; P, posterior. Scale bar, 200 μ m. **b**, Lateral semi-thin section showing PGC cluster in a late streak stage embryo stained by methylene blue, with anterior to the left. The black and white arrows indicate the incipient allantois and the chorionic membrane, respectively. White arrowheads indicate densely stained cells in the cluster. This region from early bud stage was dissected from E 7.5 embryos, and single cells were isolated for analysis. Scale bar, 50 μ m. **c**, Southern blot analysis of marker genes in single cell cDNAs. In each experiment, 20 cells were picked randomly and their cDNAs were amplified. The samples were examined sequentially for *Bmp4*, *Tnap* and *Hoxb1* expression. Experiment and clone numbers are indicated above and below the bars, respectively. An asterisk indicates PGC candidates, which are strongly positive for *Tnap* and negative for *Hoxb1*. *Gapdh* was a control for amplification efficiency. **d**, Summary of 83 single cell cDNAs analysed.

detected in the E12.5 gonadal PGCs (Fig. 3d), as demonstrated by co-labelling with anti-SSEA1 antibody (TG1), a marker for gonadal PGCs.

The relationship between *fragilis* and *stella*

As shown above, PGCs are recruited from among the *fragilis* positive community of cells. To determine the relationship between *fragilis* and *stella*, we used molecular analysis of single cell cDNAs, which is a powerful tool for the analysis of PGC specification (Figs 4 and 5). We found that PGCs consistently exhibited higher expression of *fragilis* (and *Tnap*) compared with the levels in the adjacent somatic cells. More significantly, expression of *stella* was confined to the nascent PGCs with the highest levels of *fragilis* and no *Hoxb1* expression (see below). Indeed, *stella* expression was never detected in cells where *fragilis* was absent or expressed at low levels. However, we found two cells that were negative for *stella* despite being strongly positive for *fragilis* and *Tnap*, and negative for *Hoxb1* (numbers 10 and 76): these cells may either express *stella* a little later or they may eventually fail to acquire the germ cell fate.

Analysis of *Bmp4* positive embryonic cells

Having established a clear molecular distinction between the nascent PGCs and somatic neighbours, we returned to the 57 *Bmp4* positive cells that were previously set aside (see Fig. 1d). Among these, we found 47 *Tnap* positive cells, of which four were also *stella* positive. These four cells also expressed *fragilis* at a higher level but not *Hoxb1* (Supplementary Information 1 and 2). By

contrast, of the 18 cells with lower expression of *fragilis* that we analysed, all were *Hoxb1* positive but none of them expressed *stella*. Thus, expressions of *Bmp4* and PGC specification are not mutually exclusive. Indeed, all the *Bmp4* positive/*stella* positive cells showed essentially identical molecular phenotype to *Bmp4* negative/*stella* positive cells. Thus, of the original 83 cells (Fig. 1d), 12 were *stella* positive, which gives us an estimate of about 43 nascent PGCs in the original embryonic fragment of approximately 300 cells, which is consistent with the expected number of founder PGCs.

Repression of *Hox* genes in nascent PGCs

The molecular analysis of single cells also provided additional information on the specification of PGCs (Figs 4 and 5; see also Supplementary Information 1 and 2). As indicated previously, *Oct4* (a central gene associated with pluripotency^{16–18}) showed a wider expression pattern that was not just confined to PGCs. Furthermore, both the PGCs and their somatic neighbours expressed the principal mesodermal genes, *T* (*brachyury*) and *Fgf8* (refs 29–31). Thus germ cell fate is probably imposed on cells that are initially destined towards a somatic mesodermal fate. Most importantly, however, we found that germ cell specification involves repression of the region-specific homeobox genes, *Hoxa1*, *Hoxb1*, *Lim1* and *Evx1* (refs 23, 32–34), which are all expressed in the adjacent somatic cells. As some nascent PGCs show low levels of *Hoxa1* and *Evx1* expression, this suggests that the repression of homeobox genes is an active process through which PGCs escape from a somatic cell fate and re-acquire the pluripotent state. For example, the expression of

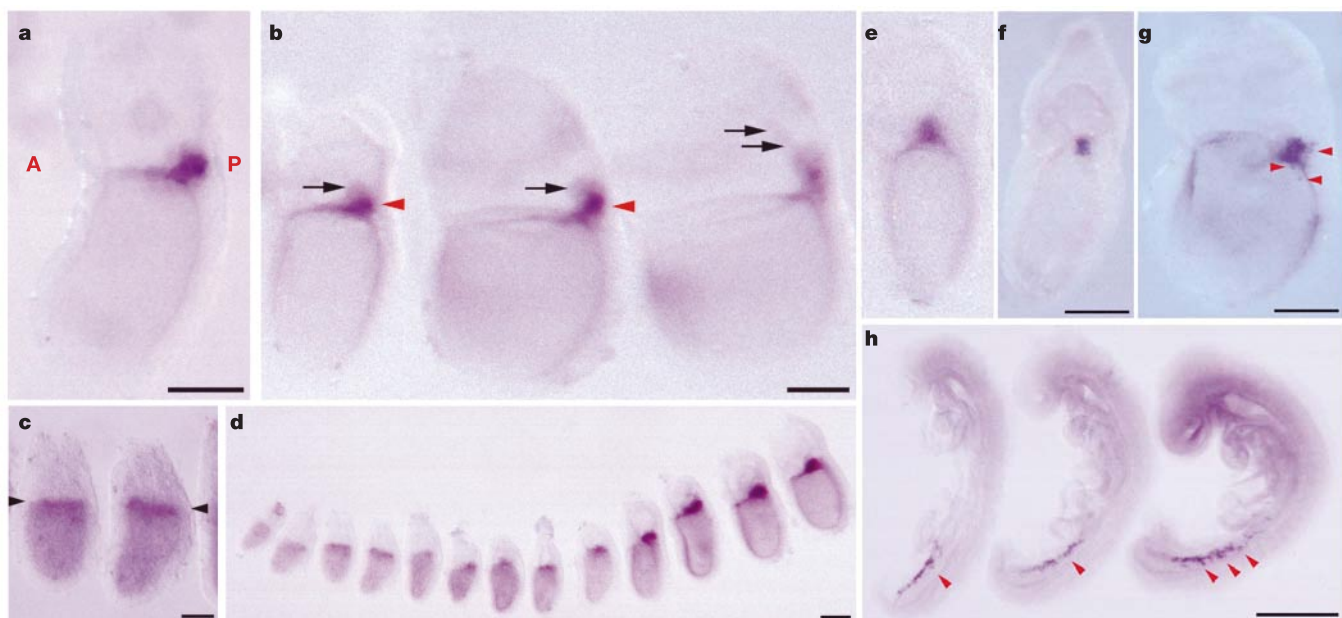


Figure 2 *fragilis* (a–e) and *stella* (f–h) expression in the germline-competent and lineage-restricted germ cells during gastrulation. **a**, Lateral view of an early bud stage embryo with expression of *fragilis*. Anterior (A) is to the left and posterior (P) to the right. Strong expression was observed specifically at the base of the incipient allantois, the location of nascent PGCs. Note a strong signal in the central region and weaker at the periphery. **b**, *fragilis* expression from mid bud (farthest left) to early head fold (farthest right) stages, as viewed from the lateral side. *fragilis* expression is strong in the centre (arrowheads)—the site of the founder PGC cluster also stained by TNAP¹⁴ (see Fig. 1a). *fragilis* expression fades at the early head fold stage (extreme right), when PGCs commence migration. Arrows indicate developing allantois where *fragilis* expression is weak or absent. **c**, Lateral views of pre-streak stage embryos (E6.25) with expression of *fragilis*. Intense signal was observed in proximal epiblast cells adjacent to the extra-embryonic ectoderm (arrowheads). **d**, *fragilis* expression from pre-streak (E6.0) (farthest left) to early bud

(farthest right) stages. The initial domain of *fragilis* expression in the most proximal epiblast followed by its movement to the posterior region during gastrulation is shown. This expression pattern matches with the observations from clonal analysis for the origin, migration and segregation of the germ cell lineage⁹. **e**, **f**, Posterior view of expression of *fragilis* at late streak stage (**e**) and a similar posterior view of a late streak stage embryo with expression of *stella* (**f**). *fragilis* expression is more extensive and slightly diffuse, compared with *stella*, which is restricted within single cells of the nascent PGC cluster. This conclusion is consistent with the molecular analysis of single cells shown in Figs 4 and 5. **g**, Posterior view of a mid bud stage embryo with migrating PGCs from the cluster expressing *stella* (arrowheads). **h**, Lateral view of E8.5 embryos with migrating PGCs (arrowheads) demonstrating expression of *stella*. Scale bars: **a**, **b**, **d–g**, 200 μm; **c**, 100 μm; **h**, 500 μm.

Evx1 in PGCs derived from the late streak stage (about 6 h earlier) is as high as that in somatic cells (data not shown), which supports our proposal for the active suppression of *Hox* genes during specification of PGCs. Therefore, a probable molecular sequence leading to PGC specification involves strong expression of *fragilis*, which always precedes *Hoxb1* repression, followed by the expression of *stella*. We also note that expression of housekeeping genes, including β -actin and *Hprt*, was consistently lower in the PGCs, which may reflect either their lengthening cell cycle time or a transient decrease of metabolic activity.

Finally, expression of the mammalian homologues of genes such as *Vasa* that are expressed in the nascent germ cells of lower organisms, were not detected in the founder PGCs in mice, although they are expressed later on^{4,35,36}. Apart from mammals, urodeles also apparently lack germ plasm. As in mice, germ cells in axolotls seem to be induced in the mesoderm³⁷, which also express mesodermal markers including *Axbra*^{37–39}. Furthermore, as in mice, expression of *Vasa* (and other genes) in axolotls also occurs during the entry of germ cells into the gonads^{38,40}. These findings emphasize that a different mechanism operates for the specification of the germ line in the absence of preformed germ plasm. Signalling molecules must induce germ cell competence in such cases.

Induction of *fragilis* by extra-embryonic ectoderm *in vitro*

Next, we examined whether *fragilis* can be induced in epiblast cells by intimate contact with the extra-embryonic ectoderm. Indeed, we detected strong expression of *fragilis* in distal epiblast cells at E6.0

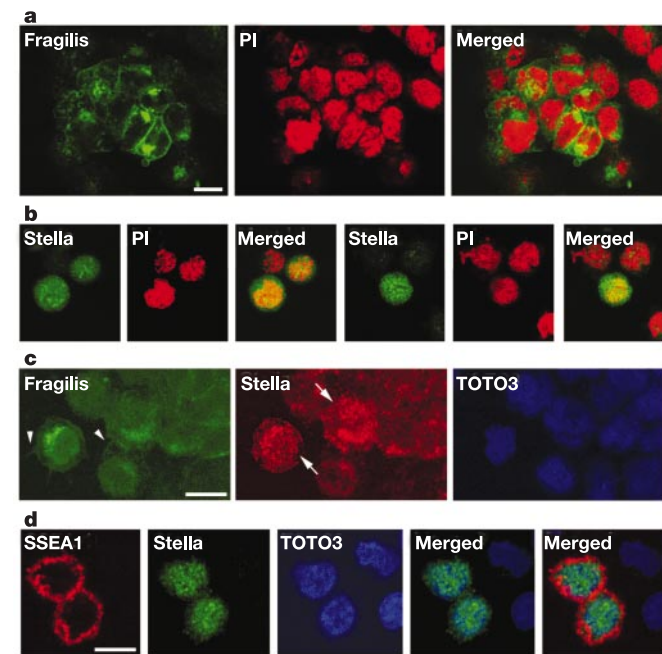


Figure 3 Localization of *fragilis* and *stella* by immunofluorescence confocal microscopy **a**, Partially dissociated cells from an early bud stage (E7.5) embryonic fragment bearing PGCs doubly stained with a rabbit anti-*fragilis* polyclonal antibody and propidium iodide (PI). Note that *fragilis* is at the cell surface. **b**, A similar fragment as in **a** stained with a rabbit anti-*stella* polyclonal antibody and propidium iodide. *Stella* is in the nucleus, and to a lesser extent in the cytoplasm. **c**, A similar fragment triple-stained with a rabbit anti-*stella* antibody, a FITC-conjugated rabbit anti-*fragilis* antibody and TOTO3. Arrowheads indicate a filopodia-like process extending from the two *fragilis* and *stella* double positive cells. **d**, Dissociated cells of E12.5 male genital ridges were triple-stained with a mouse germ-cell-specific anti-SSEA1 monoclonal antibody (TG1), an anti-*stella* polyclonal antibody and TOTO3. Note that *stella* is present in PGCs that are positive for SSEA1. Scale bars, 10 μ m.

(which do not normally give rise to PGCs) when in direct contact with the extra-embryonic ectoderm, but neither the isolated control pieces of epiblast nor extra-embryonic ectoderm alone did so (Fig. 6a). This induction was stronger closest to the site of contact and decreased progressively at the distal end.

One of the principal signals from extra-embryonic ectoderm that is implicated in PGC specification is *Bmp4* (ref. 6). To examine whether *Bmp4* affects expression of *fragilis*, we examined the expression of *fragilis* in *Bmp4*-deficient heterozygous and homozygous embryos. As shown in Fig. 6b, we detected strong *fragilis* expression in the wild-type embryos, whereas no signal was observed in the homozygous *Bmp4*^{-/-} embryos. Furthermore, expression of *fragilis* was substantially reduced in the *Bmp4* heterozygous embryos, which is in agreement with the dose-dependent effect of *Bmp4* on the number of founder PGCs, without an effect on most of the cells that contribute to somatic tissues such as the allantois⁶.

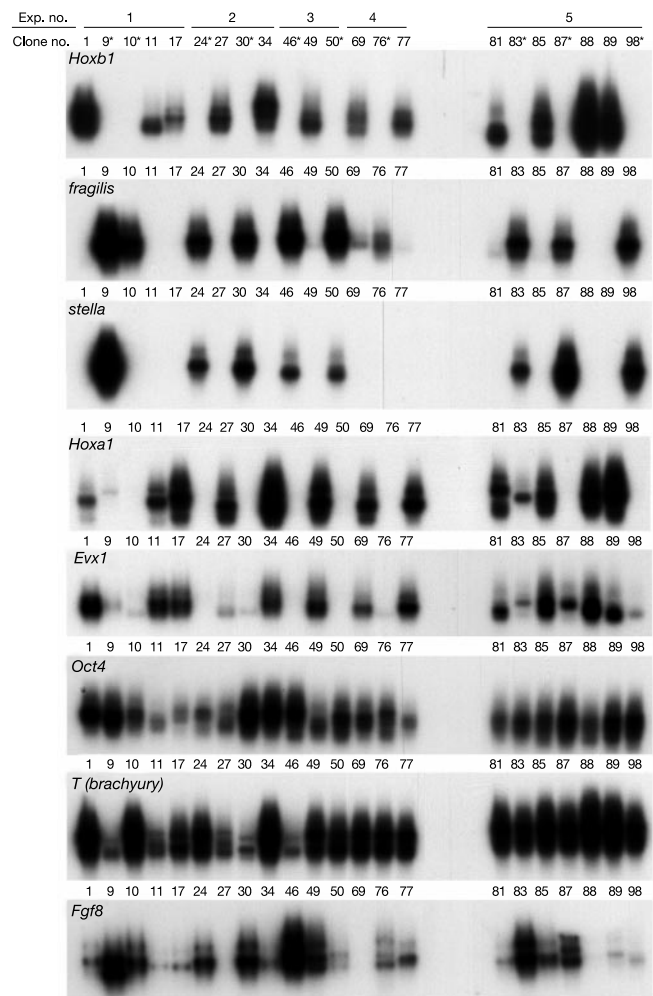


Figure 4 Molecular characterization of *Bmp4* negative single PGCs and neighbouring somatic cells. Expression of *Hoxb1*, *fragilis*, *stella*, *Hoxa1*, *Evx1*, *Oct4*, *T (brachyury)* and *Fgf8* examined by Southern blot analysis using 5 μ g of single cell cDNAs. PGC candidates, indicated by an asterisk, are strongly positive for *Tnap*, *fragilis* and *stella* but are negative for *Hoxb1* (see Fig. 1c). Note that *Hoxa1* and *Evx1* are also repressed in PGCs. Somatic cells show weak expression of *fragilis* but strong expression of *Hoxb1*. *Oct4* is expressed to a similar extent in all the cells. Note the strong expression of mesodermal markers *T (brachyury)* and *Fgf8* in the *stella* positive nascent PGCs. The expression profiles are summarized in Fig. 5.

Analysis of induced distal epiblast cells

Molecular analysis of single epiblast cells induced *in vitro* for 40–88 h by extra-embryonic ectoderm showed that 39 out of 73 cells were positive for *fragilis*, of which 15 expressed it at high levels (Fig. 7, and data not shown). Furthermore, 11 out of 39 *fragilis* positive cells showed expression of *Tnap*, among which, five were also positive for *Oct4*. Of these, 14 out of 24 *fragilis* positive cells that were also positive for *Tnap* and *Oct4*, expressed *T (brachyury)*, and 18 out of 24 were also *Bmp4* positive; a marker for the posterior extra-embryonic mesoderm. However, none of the cells showed expression of *stella*, a finding we confirmed in an additional 35 out of 68 *fragilis* positive cells that we analysed. Among these single cells, we were able to distinguish extra-embryonic ectoderm cells because they were negative for *fragilis*, *T (brachyury)* and *Oct4*, but instead showed high expression of *H19* and *Bmp8b* (ref. 10) (data not shown).

These experiments show that the distal epiblast cells induced by extra-embryonic ectoderm *in vitro* acquire germ cell competence and progress significantly towards a germ cell fate as shown by their expression of *fragilis*, *Tnap*, *Oct4* and *T (brachyury)*. Similar experiments by others have also previously shown induction of *Tnap* and *Oct4*, from which it has been suggested that extra-embryonic ectoderm can induce PGCs *in vitro*⁹; however, as we have failed to detect *stella*, it is possible that an additional signal may be required for the final stages of PGC specification *in vivo*.

The involvement of a localized second signal within the niche where PGC specification finally occurs has been postulated previously^{6,7}.

Organization and demarcation of the nascent PGC cluster

From our combined studies of the relevant region, a consistent picture of the organization of the cluster of nascent PGCs and surrounding somatic cells emerges (see Fig. 8). The expression of *fragilis* in the proximal rim of the epiblast marks the onset of germ cell competence, which is dependent on contact with the extra-embryonic ectoderm. Furthermore, the size of the *fragilis* positive cluster is apparently affected by the gene dosage of *Bmp4*, being much reduced in *Bmp4*^{+/-} embryos. This suggests that the size of the *fragilis* positive cluster is a central determinant of the number of nascent PGCs. However, the absolute numbers of *fragilis* positive cells in the cluster are not critical, because as few as 2–4 PGCs are found in *Bmp4* and *Bmp8b* mutants¹⁰. Thus, germ cells are recruited from a community of cells⁴¹ that express *fragilis*, and thus are not from a pioneer cell. We assume however that the influence of *Bmp4* would be indirect, as *fragilis*, like other family members⁴², has two strongly conserved IFN response elements upstream of its exon 1 (data not shown). *fragilis* is therefore likely to be induced by IFN, perhaps after an initial priming of the most proximal epiblast cells by *Bmp4* to become responsive to IFNs. IFNs have been

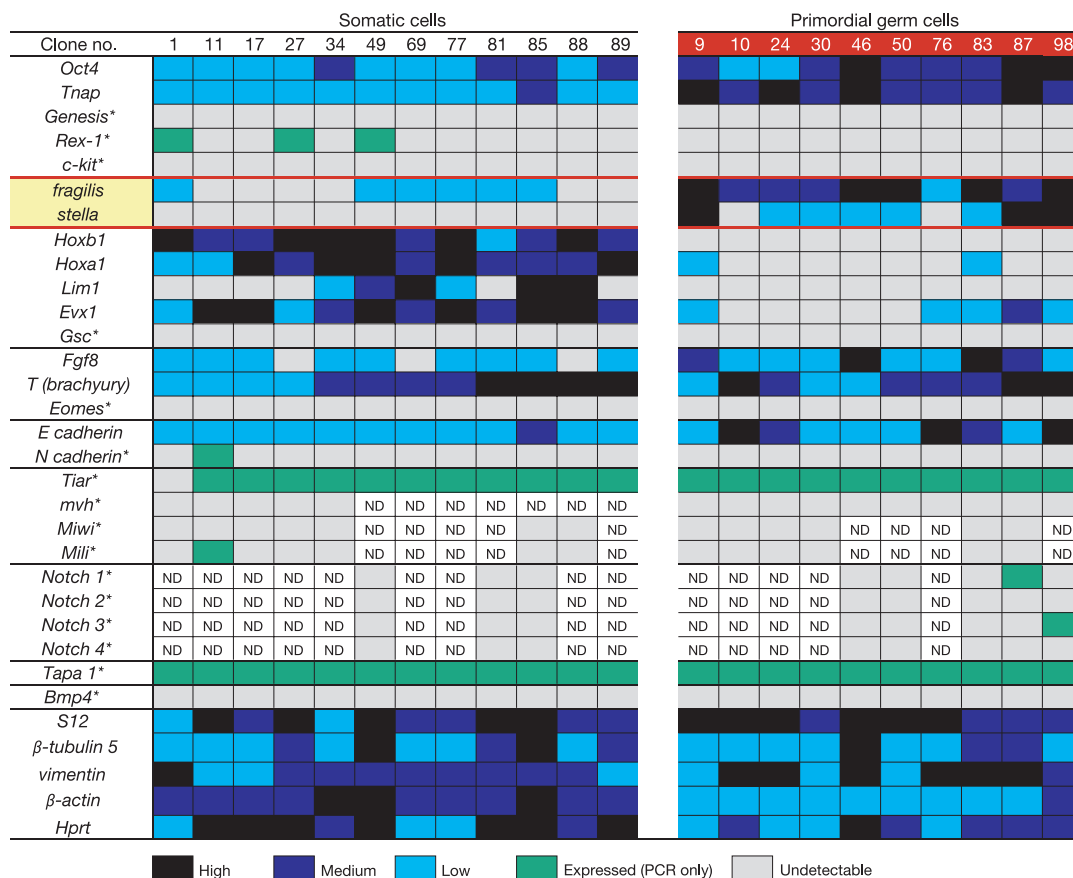


Figure 5 Gene expression profiles of *Bmp4* negative single PGCs and neighbouring somatic cells. Expression of various genes in ten individual PGCs (marked in red on the right) and the adjacent 12 individual somatic cells (left) are shown. The expression levels for each of the genes was normalized against *Gapdh*. The highest level of expression detected was assigned a value of 100. Cells with a value higher than 66 were classified as

high (black), between 33 and 66 as medium (dark blue), and lower than 33 as low (blue). Thus, the relative levels of expression for each gene in different individual cells is shown. Expression of some of the genes, indicated by an asterisk, was examined by PCR only, and therefore their relative levels are not indicated. ND, not determined.

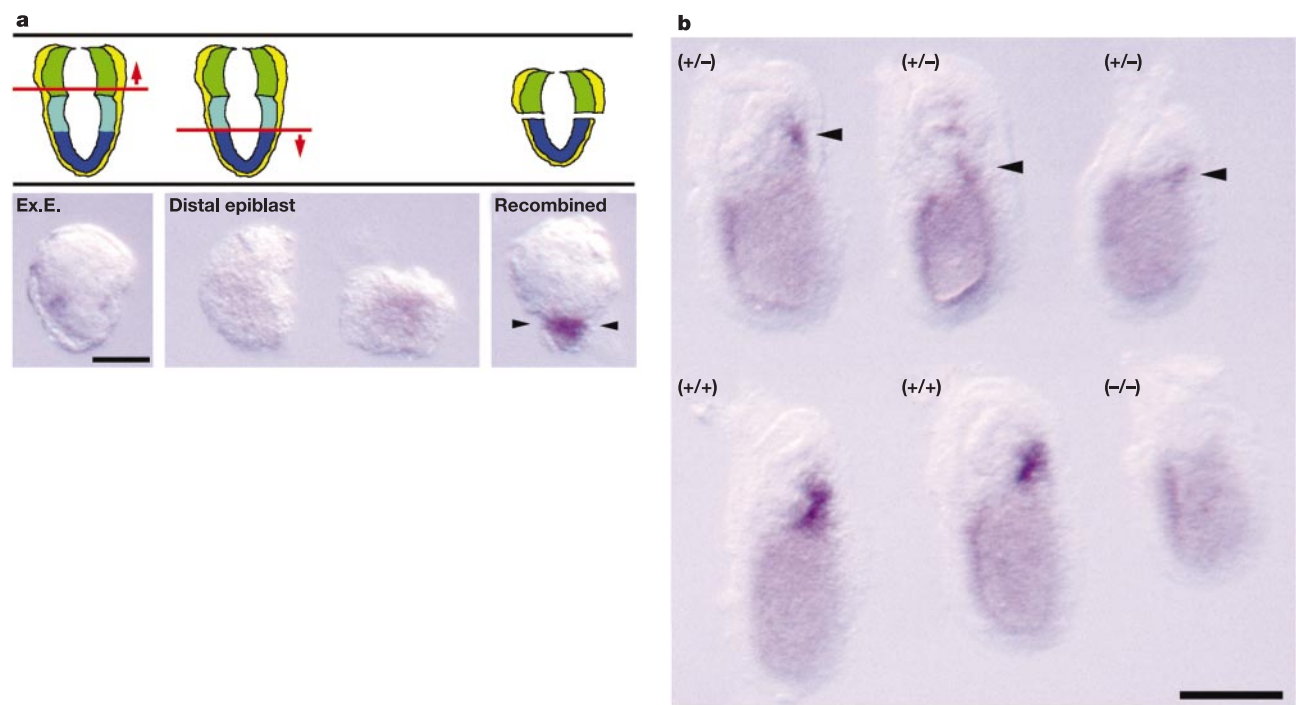


Figure 6 *fragilis* expression is inducible by extra-embryonic tissues and is dose dependent on Bmp4. **a**, Schematic representation of the embryo recombination experiment (top panel). E6.0 embryos (green, extra-embryonic ectoderm; pale blue, proximal epiblast; blue, distal epiblast; yellow, visceral endoderm) were cut with glass needles along the red lines indicated, and cultured. Fusion of distal epiblast and extra-embryonic ectoderm (Ex.E.) was brought about by placing the two fragments with their cut faces apposed in a drop of medium (approximately 15 μl) in a small

depression well, covered with mineral oil. The bottom panel shows clear *fragilis* expression at the border between distal epiblast and extra-embryonic ectoderm (arrowhead), only in the recombined embryo. Scale bar, 100 μm. (See Fig. 7 for an examination of these cells in detail.) **b**, Expression of *fragilis* in wild-type (+/+), *Bmp4* heterozygous (+/-) and homozygous (-/-) embryos. Arrowheads indicate smaller clusters of *fragilis* positive cells in the heterozygous embryos. Scale bar, 200 μm.

reported during gastrulation, which is consistent with this suggestion^{43,44}.

Proposed roles of *fragilis* and *stella* in PGC specification

Fragilis as a typical IFN-inducible cell surface protein, probably shares certain properties common to all of these family members^{24,26,27}. The acute but transient upregulation of *fragilis* is itself consistent with the kinetics of IFN-inducible genes that can increase by up to 40-fold within 1 h, and decline quickly after IFN withdrawal²⁵. This *fragilis* positive assembly of cells could correspond to about 100 TNAP positive cells^{5,14}, which is larger than the number of

stella positive cells. According to our estimates, the *stella* positive cluster in the 129/SvEv mouse strain consists of approximately 36–43 cells, which is close to the expected 45 nascent PGCs. The *fragilis* positive cells probably form a community of cells through homotypic adhesion^{26,27}, from which the founder PGCs are recruited, thus demarcating them from most of the cells destined for somatic tissues. These IFN-inducible cell surface proteins are capable of transduction of antiproliferative signals²⁴, which is a probable mechanism by which the cell cycle time in the nascent PGCs increases from 6 to 16 h, while the somatic cells continue to divide rapidly⁵.

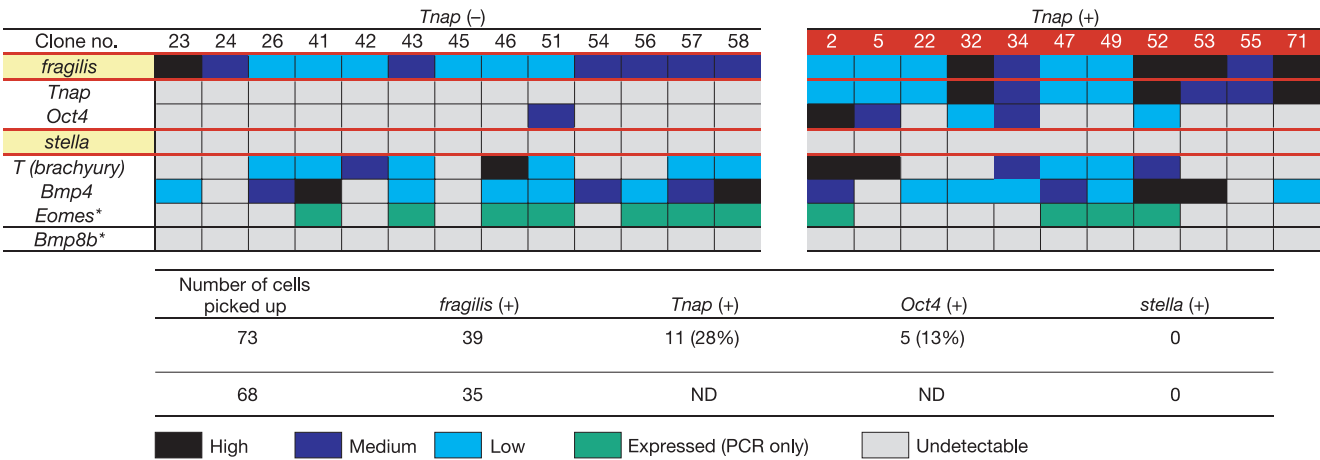


Figure 7 Gene expression profiles of distal epiblast cells induced by extra-embryonic ectoderm *in vitro*. The levels of expression were determined as described for Fig. 5. ND, not determined.

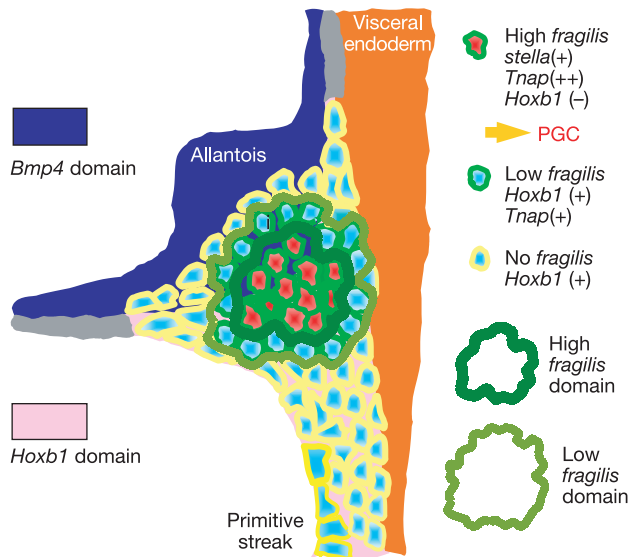


Figure 8 Organization of the niche for PGC specification. PGCs consist of both *Bmp4* positive and negative cells at the proximal and distal ends of the cluster, respectively. PGCs exhibit high levels of *fragilis* and *Tnap*, and more crucially, they express *stella* but not *Hoxb1*. This core PGC cluster is surrounded by somatic cells, with lower or undetectable levels of *fragilis* and *Tnap*, but they express region-specific homeobox genes, including *Hoxb1*. The proximal *Bmp4* positive cells correspond to the cells that accumulate in the bud region even in the absence of *Bmp4* expression⁶, whereas the distal *Bmp4* negative cells probably constitute a separate lineage or start expressing this gene later.

The induction of *fragilis* in epiblast cells may not by itself be sufficient for the expression of *stella*, as shown by our *in vitro* studies—induction may require a specific signal thought to be within the niche, for PGC specification *in vivo*^{6,7}. This signal could be a specific ligand that binds to *fragilis* during the specification of germ cell fate. Once nascent PGCs are established, expression of *fragilis* is diminished by E8.0 (Fig. 2b), thus freeing the PGCs from homotypic adhesion for their migration into the genital ridge^{4,45}.

fragilis must have other functions, as it is apparently expressed elsewhere in developing embryos. In this context, we also note *fragilis* expression in pluripotent ES and embryonic germ cells (data not shown), where it may have a role in the propagation of the pluripotent state. The role of *stella* may in part be regulated by its potential to shuttle between the nucleus and cytoplasm. We have observed, for example, that overexpression of *stella* in some somatic cells causes the protein to be retained in the cytoplasm and not in the nucleus, as is predominantly the case in PGCs (data not shown).

A particularly critical event involved in the specification of PGCs is repression of the region-specific homeobox genes, by which nascent PGCs escape from the somatic cell fate. As the expression of *stella* is most intimately connected with the generation of PGCs, this gene is a candidate for either initiating or maintaining repression of *Hox* genes in PGCs. The detection of *stella* in the oocyte and through pre-implantation development (B. Payer *et al.*, unpublished data; see also ref. 46) suggests that it may serve a critical role during all the phases of totipotent/pluripotent states in mice.

Conclusion

With the single cell analysis of PGCs and neighbouring somatic cells that share common ancestry, we have established a systematic investigation of the genetic basis for germ cell specification in mice. Here we propose that *fragilis*, an interferon-inducible trans-

membrane protein, promotes germ cell competence and homotypic association to demarcate putative germ cells from their somatic neighbours. Such involvement of an IFN-inducible gene may provide a new developmental model of wide applicability. The rapid and transient homotypic adhesion of cells through such an IFN-responsive gene is unlike conventional adhesion mechanisms²⁶. We propose that this is a mechanism for developmental decisions and demarcation of a subset of cells especially during dynamic morphogenetic movements in the embryo. The nascent population of PGCs is recruited from a community of cells demarcated by high expression of *fragilis*, which is followed by the expression of *stella*. All of these cells are initially destined for a mesodermal fate with high expression of *T* (*brachyury*) and *Fgf8*. However, with the repression or failure of derepression of the region-specific homeobox genes, the founder PGCs escape from a somatic cell fate and reacquire pluripotency. Once the specification of germ cell fate is accomplished, expression of *fragilis* quickly wanes, but *stella* continues in the germ cell lineage thereafter. Indeed, *stella* may be crucial for the totipotent/pluripotent states, beginning with the oocyte and continuing through pre-implantation development. □

Methods

Histology

We performed alkaline phosphatase staining as described⁶. For semi-thin section analysis, gastrulating embryos were fixed in 4% glutaraldehyde in 0.1 M PIPES buffer, pH 7.2, for 2 h at room temperature and processed as described⁴⁷.

Preparation and screening of single cell cDNA library

Early bud stage 129/SvEv embryos were isolated at E7.5 in DMEM with 10% fetal calf serum buffered with 25 mM HEPES, pH 7.4, at room temperature. Fragments bearing PGCs were dissected out using glass needles and dissociated into single cells by brief treatment with trypsin, EDTA and PBS, followed by gentle pipetting. Single cells were picked by mouth pipettes and their cDNAs were amplified and screened essentially as described previously^{20,21}.

In situ hybridization and immunohistochemistry

In situ hybridization was performed as described^{48,49}. Rabbit polyclonal anti-*fragilis* and anti-*stella* antibodies were raised against two peptides (ASGGQPPNYRIKEEYE and RDRKMVGVDVTGAQAYA, and MEPPSEKVDPMKDPET and CHYQRWDPSSENK IGKN, respectively) and affinity purified using the same peptides. Single cell preparations from E7.5 embryonic fragments bearing PGCs and E12.5 genital ridges were fixed in 4% paraformaldehyde in PBS for 10 min at room temperature and processed for immunohistochemistry as described⁴⁷. For double staining, dissociated embryonic cells were first stained by a rabbit polyclonal anti-*stella* antibody and after extensive washing, labelled with anti-rabbit immunoglobulin-γ antibody coupled with Alexa 568 (Molecular Probe). These cells were again extensively washed and incubated with an anti-*fragilis* antibody conjugated with fluorescein isothiocyanate (FITC) and enhanced by a mouse anti-FITC antibody coupled with Alexa 468 (Molecular Probe).

Embryo recombination culture experiment

Embryos from 129/SvEv female mice were isolated at E6.0, and pieces of extra-embryonic tissues and distal epiblasts were cut from the embryos with glass needles. For recombination, a piece of distal epiblast was put in a small depression well with its cut face towards the cut face of an extra-embryonic ectoderm fragment. They were cultured in a small drop of medium (50% DMEM plus 50% rat serum or 85% DMEM plus 15% FCS) covered with mineral oil. Some of them fused after several hours of culture at 37 °C in 5% CO₂ in air. After 36 h, the expression of *fragilis* was examined by *in situ* hybridization. For the preparation of single cell cDNAs from recombined embryos after 40–88 h of culture, the junctional part was cut to dissociate and isolate cells induced by extra-embryonic ectoderms.

Bmp4-deficient embryos

A strain of *Bmp4*^{tm1bH} mice⁵⁰, on a C57BL/6/CBA mixed background, was the gift of K. Lawson. Heterozygous mice were crossed to obtain homozygous embryos at E7.0. Embryos were genotyped after *in situ* hybridization.

Received 9 April; accepted 13 June 2002; doi:10.1038/nature00927.

1. Weismann, A. *Das Keimplasma. Eine Theorie der Vererbung* (Gustav Fischer, Jena, 1892).
2. Eddy, E. M. Germ plasma and the differentiation of the germ cell line. *Int. Rev. Cytol.* **43**, 229–280 (1975).
3. Seydoux, G. & Strome, S. Launching the germline in *Caenorhabditis elegans*: regulation of gene expression in early germ cells. *Development* **126**, 3275–3283 (1999).
4. Wylie, C. Germ cells. *Cell* **96**, 165–174 (1999).
5. Lawson, K. A. & Hage, W. J. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found. Symp.* **182**, 68–84 (1994).
6. Lawson, K. A. *et al.* *Bmp4* is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev.* **13**, 424–436 (1999).

7. McLaren, A. Signaling for germ cells. *Genes Dev.* **13**, 373–376 (1999).
8. Tam, P. P. & Zhou, S. X. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev. Biol.* **178**, 124–132 (1996).
9. Yoshimizu, T., Obinata, M. & Matsui, Y. Stage-specific tissue and cell interactions play key roles in mouse germ cell specification. *Development* **128**, 481–490 (2001).
10. Ying, Y., Liu, X. M., Marble, A., Lawson, K. A. & Zhao, G. Q. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol. Endocrinol.* **14**, 1053–1063 (2000).
11. Ying, Y., Qi, X. & Zhao, G. Q. Induction of primordial germ cells from murine epiblasts by synergistic action of BMP4 and BMP8B signaling pathways. *Proc. Natl Acad. Sci. USA* **98**, 7858–7862 (2001).
12. Ying, Y. & Zhao, G. Q. Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. *Dev. Biol.* **232**, 484–492 (2001).
13. Chiquoine, A. D. The identification, origin and migration of the primordial germ cells in the mouse embryo. *Anat. Rec.* **118**, 135–146 (1954).
14. Ginsburg, M., Snow, M. H. & McLaren, A. Primordial germ cells in the mouse embryo during gastrulation. *Development* **110**, 521–528 (1990).
15. MacGregor, G. R., Zambrowicz, B. P. & Soriano, P. Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during mouse embryogenesis but is not required for migration of primordial germ cells. *Development* **121**, 1487–1496 (1995).
16. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).
17. Pesce, M., Gross, M. K. & Scholer, H. R. In line with our ancestors: Oct-4 and the mammalian germ. *Bioessays* **20**, 722–732 (1998).
18. Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894 (1996).
19. Downs, K. M. & Davies, T. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* **118**, 1255–1266 (1993).
20. Brady, G. & Iscove, N. N. Construction of cDNA libraries from single cells. *Methods Enzymol.* **225**, 611–623 (1993).
21. Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195–206 (1995).
22. Tanabe, Y., William, C. & Jessell, T. M. Specification of motor neuron identity by the MNR2 homeodomain protein. *Cell* **95**, 67–80 (1998).
23. Frohman, M. A., Boyle, M. & Martin, G. R. Isolation of the mouse Hox-2.9 gene; analysis of embryonic expression suggests that positional information along the anterior-posterior axis is specified by mesoderm. *Development* **110**, 589–607 (1990).
24. Deblandre, G. A. *et al.* Expression cloning of an interferon-inducible 17-kDa membrane protein implicated in the control of cell growth. *J. Biol. Chem.* **270**, 23860–23866 (1995).
25. Friedman, R. L., Manly, S. P., McMahon, M., Kerr, I. M. & Stark, G. R. Transcriptional and posttranscriptional regulation of interferon-induced gene expression in human cells. *Cell* **38**, 745–755 (1984).
26. Evans, S. S., Collea, R. P., Leasure, J. A. & Lee, D. B. IFN- α induces homotypic adhesion and Leu-13 expression in human B lymphoid cells. *J. Immunol.* **150**, 736–747 (1993).
27. Evans, S. S., Lee, D. B., Han, T., Tomasi, T. B. & Evans, R. L. Monoclonal antibody to the interferon-inducible protein Leu-13 triggers aggregation and inhibits proliferation of leukemic B cells. *Blood* **76**, 2583–2593 (1990).
28. Aravind, L. & Koonin, E. V. SAP—a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem. Sci.* **25**, 112–114 (2000).
29. Crossley, P. H. & Martin, G. R. The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121**, 439–451 (1995).
30. Herrmann, B. G. Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. *Development* **113**, 913–917 (1991).
31. Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. & Lehrach, H. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617–622 (1990).
32. Barnes, J. D., Crosby, J. L., Jones, C. M., Wright, C. V. & Hogan, B. L. Embryonic expression of Lim-1, the mouse homolog of *Xenopus* Xlim-1, suggests a role in lateral mesoderm differentiation and neurogenesis. *Dev. Biol.* **161**, 168–178 (1994).
33. Bastian, H. & Gruss, P. A murine even-skipped homologue, Evx 1, is expressed during early embryogenesis and neurogenesis in a biphasic manner. *EMBO J.* **9**, 1839–1852 (1990).
34. Fujii, T. *et al.* Expression patterns of the murine LIM class homeobox gene lim1 in the developing brain and excretory system. *Dev. Dyn.* **199**, 73–83 (1994).
35. Kuramochi-Miyagawa, S. *et al.* Two mouse piwi-related genes: miwi and mili. *Mech. Dev.* **108**, 121–133 (2001).
36. Fujiwara, Y. *et al.* Isolation of a DEAD-family protein gene that encodes a murine homolog of *Drosophila* vasa and its specific expression in germ cell lineage. *Proc. Natl Acad. Sci. USA* **91**, 12258–12262 (1994).
37. Nieuwkoop, P. D. & Sataura, L. A. *Primordial Germ Cells in the Chordates* (Cambridge Univ. Press, Cambridge, 1979).
38. Johnson, A. D., Bachvarova, R. F., Drum, M. & Masi, T. Expression of axolotl dazl RNA, a marker of germ plasm: widespread maternal RNA and onset of expression in germ cells approaching the gonad. *Dev. Biol.* **234**, 402–415 (2001).
39. Johnson, A. D., Bachvarova, R. F., Masi, T. & Drum, M. in *Germ Cells: Cold Spring Harbor Lab. Meeting* 61 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2000).
40. Toyooka, Y. *et al.* Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mech. Dev.* **93**, 139–149 (2000).
41. Gurdon, J. B., Kato, K. & Lemaire, P. The community effect, dorsalization and mesoderm induction. *Curr. Opin. Genet. Dev.* **3**, 662–667 (1993).
42. Reid, L. E. *et al.* A single DNA response element can confer inducibility by both α - and γ -interferons. *Proc. Natl Acad. Sci. USA* **86**, 840–844 (1989).
43. Barlow, D. P., Randle, B. J. & Burke, D. C. Interferon synthesis in the early post-implantation mouse embryo. *Differentiation* **27**, 229–235 (1984).
44. Kita, M. *et al.* Expression of cytokines and interferon-related genes in the mouse embryo. *C.R. Seances Soc. Biol. Fil.* **188**, 593–600 (1994).
45. Gomperts, M., Garcia-Castro, M., Wylie, C. & Heasman, J. Interactions between primordial germ cells play a role in their migration in mouse embryos. *Development* **120**, 135–141 (1994).
46. Sato, M. *et al.* Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells. *Mech. Dev.* **113**, 91–94 (2002).
47. Saitou, M. *et al.* Occludin-deficient embryonic stem cells can differentiate into polarized epithelial cells bearing tight junctions. *J. Cell Biol.* **141**, 397–408 (1998).
48. Henrique, D. *et al.* Expression of a Delta homologue in prospective neurons in the chick. *Nature* **375**, 787–790 (1995).
49. Wilkinson, D. G. & Nieto, M. A. Detection of messenger RNA by *in situ* hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225**, 361–373 (1993).
50. Winnier, G., Blessing, M., Labosky, P. A. & Hogan, B. L. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev.* **9**, 2105–2116 (1995).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank K. Lawson for providing unpublished information and for discussions, and A. McLaren for critical comments during the course of this study. We also thank Y. Tanabe and T. Yamamoto for helpful information concerning the construction of single cell cDNAs. We are grateful to the Wellcome Trust for support through a Wellcome Travelling Fellowship to M.S. and for a Programme Grant to M.A.S.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.A.S. (e-mail: as10021@mole.bio.cam.ac.uk).

Hot bubbles from active galactic nuclei as a heat source in cooling-flow clusters

Marcus Brüggen* & Christian R. Kaiser†

* International University Bremen, Campus Ring 1, 28759 Bremen, Germany

† Department of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK

Hot, X-ray-emitting plasma permeates clusters of galaxies. The X-ray surface brightness often shows a peak near the centre of the cluster that is coincident with a drop in the entropy of the gas. This has been taken as evidence for a 'cooling flow', where the gas

cools by radiating away its energy, and then falls to the centre¹. Searches for this cool gas have revealed significantly less than predicted², indicating that the mass deposition rate is much lower than expected. Most clusters with cooling flows, however, also host an active galactic nucleus at their centres³. These active galactic nuclei can inflate large bubbles of hot plasma that subsequently rise through the cluster 'atmosphere', thus stirring the cooling gas^{4,5} and adding energy. Here we report highly resolved hydrodynamic simulations which show that buoyant bubbles increase the cooling time in the inner regions of clusters and significantly reduce the deposition of cold gas.

The explanation of the absence of cold gas in the gaseous atmospheres of cooling-flow clusters strongly suggests the presence of a source of heat in their centres that is at least periodically active. Heating the gas at the cluster centre increases its entropy and thus its radiative cooling time. However, not all heating processes can

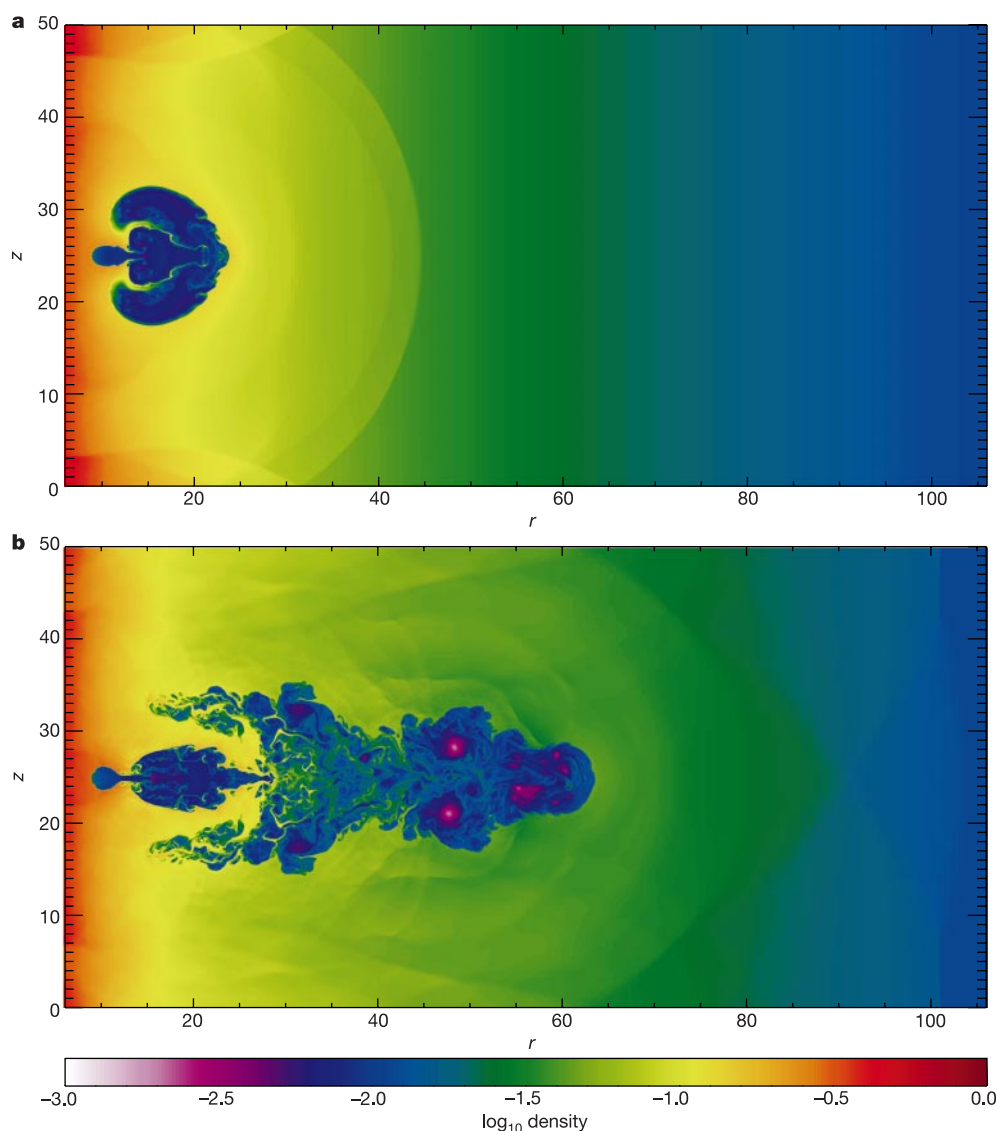


Figure 1 Snapshots of the density at two different times in our simulation (run 2). **a, b**, Contour plots showing the density after 60 Myr (**a**) and 120 Myr (**b**). The scale is logarithmic, and unit density corresponds to $3.4 \times 10^{-25} \text{ g cm}^{-3}$. The computational domain covered a rectangular region 50 kpc in width (z) and 100 kpc in height (r). The units on the axes are kpc. The injection region had a radius of 1 kpc, and the rate of energy input was $L = 4.1 \times 10^{43} \text{ erg s}^{-1}$. The energy input was switched off after 229 Myr. The temperature of the injected fluid was 100 times greater than the temperature of the ambient fluid at the same location. All simulations were run on 27 processors on an SGI

Origin 3000. We have also performed a run (run 1) with a lower rate of energy injection ($L = 2.1 \times 10^{42} \text{ erg s}^{-1}$) where the energy input was switched off after 334 Myr. At all boundaries outflow conditions were chosen. The gravitational potential was taken as static in order to save solving Poisson's equation at each time step. The gas was treated as a single fluid with a polytropic equation of state, and magnetic fields have been ignored. Radiative cooling was neglected, as the cooling time in the central regions of the cluster is still longer than the total simulated time. Numerical experiments have shown that the neglect of cooling does not influence the large-scale dynamics of the fluid.

explain the observations—for example, fully convective heating of the cluster gas would lead to the entropy being highest in the cluster centre, opposite to what is observed².

It has been suggested that the removal of cold gas and the feedback of energy through star formation and related supernovae can explain some of the observed properties of clusters⁶. Although this is a promising model for the initial formation of clusters within the framework of hierarchical structure formation in the Universe, the level of star formation found in detailed investigations of present-day cooling-flow clusters is insufficient to explain the mass deposition rates inferred from X-ray observations². Moreover, the amount of star formation during the evolution of clusters that is necessary to explain the distribution of gas observed in them today is difficult to reconcile with the observed abundances of heavy elements⁷.

Radio-loud active galactic nuclei (AGN) drive strong outflows in the form of jets that inflate bubbles or lobes. The lobes are filled with hot plasma, and can heat the cluster gas in various ways. In the case of very energetic jets, the expansion of the lobes will be supersonic, and the resulting strong shock will heat and compress the gas^{8–10}. However, there is mounting evidence that weaker jets, which are presumably much more common, do not lead to efficient shock heating¹¹. Nevertheless, they still produce pockets of low-density gas

that rise in the cluster atmosphere owing to buoyancy. As the timescale for the rise of these bubbles is long compared to the lifetime of the AGN and of the synchrotron emitting particles, we expect to observe only very few rising bubbles in the radio band¹². A substantial number of bubbles have been observed in a number of clusters as depressions in the X-ray surface brightness, sometimes associated with a low level of remnant radio emission^{8,13}. Moreover, it has been shown that a lot of energy in the form of bubbles may reside in clusters, as the bubbles are difficult to detect¹⁴. The buoyant rise and the subsequent mixing of high-entropy gas with the cluster material can lead to a significant re-structuring of the inner regions of a cooling flow. However, both the rise and the mixing are complex processes that are difficult to treat on the basis of analytical estimates alone.

We therefore performed highly resolved two-dimensional hydrodynamic simulations of buoyant gas in a typical cluster environment similar to previous simulations of lower resolution^{14,15}. These simulations were performed using the FLASH code¹⁶, which uses an explicit piece-wise parabolic method to solve the equations of compressible hydrodynamics. Only two-dimensional simulations were performed because our primary aim was to achieve a very high spatial resolution in order to minimize the effects of numerical diffusivity that have hampered previous simulations. This was also aided by the use of an adaptive grid that places resolution elements only where they are needed, such that an effective resolution of 4,000 times 2,000 elements could be achieved. Experience has shown that large-scale transport properties are not very different in two-

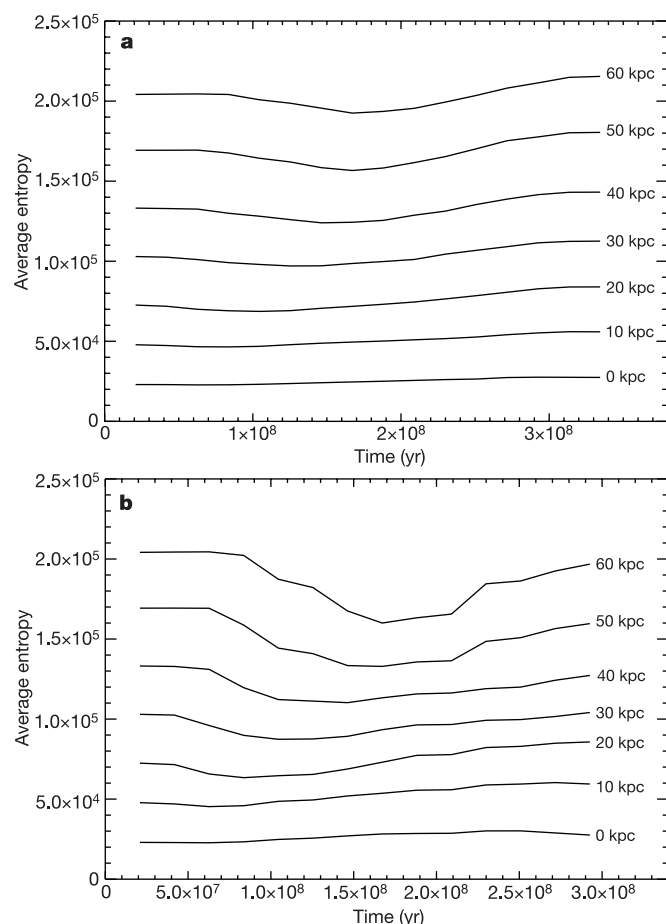


Figure 2 Average specific entropy of the ambient material in units of $3.1 \times 10^{27} \text{ erg cm}^{-2} \text{ g}^{-5/3}$ at different heights. Material with entropy below a suitably chosen entropy threshold can be identified as initially ambient material¹⁹, which enables us to study the temperature and entropy evolution of the original cluster gas alone. For this purpose, we divided the computational domain into 10 horizontal slices of 10 kpc thickness, and computed averages of physical quantities in each slice. Each curve corresponds to a separate slice, which is labelled with the height of its lower boundary. **a**, Run 1; **b**, Run 2.

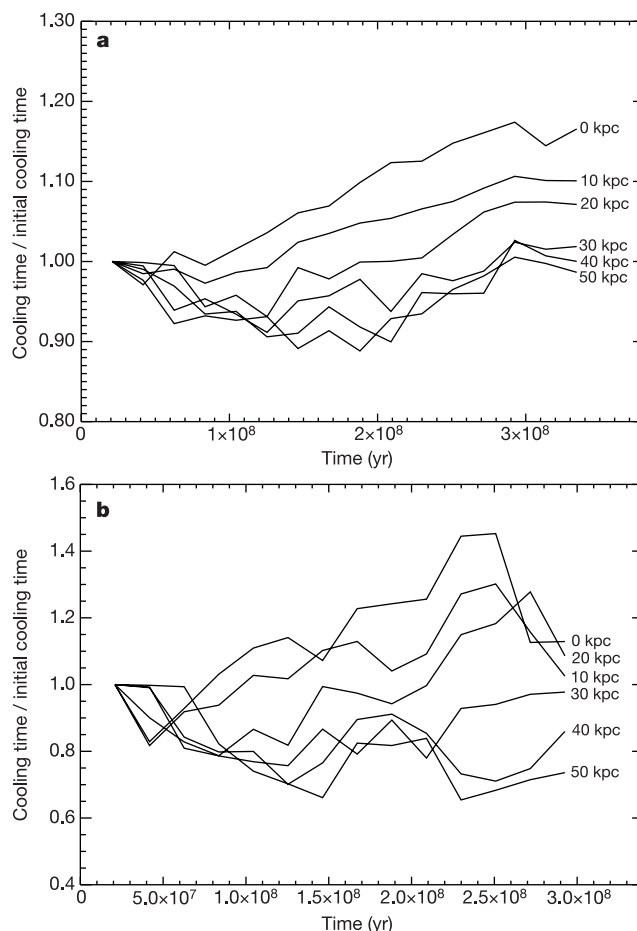


Figure 3 Cooling time of the ambient material averaged over horizontal slices. Each curve corresponds to a separate slice, which is labelled with the height of its lower boundary. For each slice, the cooling time is normalized to its value at the start of the simulation. **a**, Run 1; **b**, Run 2.

dimensional and three-dimensional simulations, so we expect our conclusions to hold even for the three-dimensional reality.

The initial distributions of mass and temperature were modelled on the well-studied Virgo cluster¹⁷, whereas the initial gas density profile was determined assuming hydrostatic equilibrium. Cosmological simulations as well as observations suggest that the gas in many clusters may not be in an equilibrium state. However, gas flows within a cluster of low Mach numbers will be unable to impede the rise of the bubble, and are more likely to increase than to decrease the mixing. On the other hand, clusters that are far from hydrostatic equilibrium are generally not expected to host cooling flows. To mimic the energy input by a jet, hot, buoyant gas was injected continuously into a spherical region offset from the gravitational centre with zero initial velocity. Figure 1 shows logarithmic contour plots of the density at 60 Myr (Fig. 1a) and 120 Myr (Fig. 1b) after the start of the energy injection (run 2). We have also performed a run (run 1) with a lower rate of energy injection. In both cases, we can see how the buoyant plume rises upwards while Kelvin–Helmholtz and secondary Rayleigh–Taylor instabilities form tori and eddies thus mixing the bubble gas with the colder cluster gas.

The horizontally averaged, specific entropy as a function of time is shown in Fig. 2. Note that the entropy at lower radii is increased at the expense of the entropy at larger radii. As all motions at this stage are subsonic and there are no shocks, the increased entropy at the cluster centre comes entirely from mixing. Shortly after the bubble has crossed a particular layer, the entropy starts to decrease, because lower-entropy gas from closer to the cluster centre is being uplifted in its wake. Then the entropy rises again, because the lower-entropy material continues to be uplifted, and higher-entropy material is being funnelled down to make place for the uplifted gas. This leads to a net increase in entropy in the lower regions. The increase in entropy in the bottom 20 kpc of the cluster over the course of the simulation is about 17.5% in run 1 and 22% in run 2. This implies that mixing is almost as effective in run 1 as in run 2, even though 13 times as much energy is injected in the latter. In both runs, the mixing process is gentle enough not to reverse the entropy gradient. In run 2, the entropy in the bottom layers starts to decrease again after the jet has been switched off, and low-entropy material starts to flow back to the centre to re-establish hydrostatic equilibrium.

Figure 3 shows the changes of the radiative cooling time of the cluster gas in the horizontal slices. Initially, the cooling time in the outer regions decreases as the rising bubble compresses the gas in front of it. However, the cooling time in the central regions increases as high-entropy material from larger radii is being mixed into the centre. Meanwhile, the low-entropy gas lifted from the cluster centre expands adiabatically, as a result of which its cooling time increases as well. In run 2, the cooling time in the bottom slice went up by as much as 40%. Note that this result applies to the original cluster gas alone. When the gas in the hot bubbles is taken also into account, the cooling time increases even more. In the cooling-flow model, mass deposition is greatest in the core, where the cooling time drops sharply. The mass that condenses out in the cores ceases to provide pressure support, and thus causes a slow inflow. Therefore, the cooling flow can be delayed significantly if the cooling of gas in the central regions can be halted.

We have found that the average temperature in the core goes up; this is because hotter material from outer regions is mixed into the core. For the parameters chosen here, the rate of increase in temperature is very close to the rate of cooling by bremsstrahlung emission. To be more precise, we have computed the ratio between the heating time ($t_{\text{heat}} = T/(dT/dt)$, where the temperature T and its time derivative are averaged over concentric annuli about the centre) and the radiative cooling time. Within the central 20 kpc this ratio varies between 1.02 and 1.15, which means that the heating is able to match the radiative cooling very closely. This leads us to conclude that the decrease of the cooling time and the mixing of

hotter material into the core both cause a significant decrease in the mass deposition rate compared to the rate predicted by the quasi-static cooling-flow model.

Clearly, the lifetime of the activity of the central AGN is short compared to the evolutionary timescale of the cluster gas. Therefore, once the AGN stops supplying energy to the buoyant bubbles, the cluster gas will settle down once again and a full cooling flow may be re-established. The cooling gas flowing to the very centre of the cluster may then trigger a further active phase of the AGN. Thus a self-regulating process—with cooling periods alternating with brief bursts of AGN activity—may be established^{18,19}. It has been found that 71% of all cD galaxies at the centres of cooling-flow clusters show evidence of radio activity³. If all AGN at the centres of cooling flows go through activity cycles of constant length, t_{on} , separated by periods of quiescence, again of constant length, t_{off} , then these observations imply that $t_{\text{off}} \approx 0.41 t_{\text{on}}$. Figures 2b and 3b indicate that the simulated cluster (run 2, $t_{\text{on}} = 239$ Myr) starts to relax back to its original state, and at time $t_{\text{on}} + t_{\text{off}} = 337$ Myr may well get close to a configuration similar to that at which the AGN activity started. Therefore, the cluster gas may never evolve beyond a state conducive to the start of AGN activity. This may explain the absence of gas with a temperature below 1 keV in clusters. The high effective resolution and the accuracy of the numerical scheme used here have made it possible to make firm conclusions about the heating and mixing of the intra-cluster medium. □

Received 28 November 2001; accepted 21 May 2002; doi:10.1038/nature00857.

1. Fabian, A. C. Cooling flows in clusters of galaxies. *Annu. Rev. Astron. Astrophys.* **32**, 277–318 (1994).
2. Böhringer, H., Matsushita, K., Churazov, E., Ikebe, Y. & Chen, Y. The new emerging model for the structure of cooling cores in clusters of galaxies. *Astron. Astrophys.* (submitted); preprint astro-ph/0111112 at (<http://xxx.lanl.gov>) (2001).
3. Burns, J. O. The radio properties of cD galaxies in Abell clusters. I – an X-ray selected sample. *Astron. J.* **99**, 14–30 (1990).
4. Churazov, E., Forman, W., Jones, C. & Böhringer, H. Asymmetric, arc minute scale structures around NGC 1275. *Astron. Astrophys.* **356**, 788–794 (2000).
5. Churazov, E., Brüggemann, M., Kaiser, C. R., Böhringer, H. & Forman, W. Evolution of buoyant bubbles in M87. *Astrophys. J.* **554**, 261–294 (2001).
6. Voit, G. M. & Bryan, G. L. Regulation of the X-ray luminosity of clusters of galaxies by cooling and supernova feedback. *Nature* **414**, 425–427 (2001).
7. Wu, K. K. S., Fabian, A. C. & Nulsen, P. E. J. Non-gravitational heating in the hierarchical formation of X-ray clusters. *Mon. Not. R. Astron. Soc.* **318**, 889–912 (2000).
8. Heinz, S., Reynolds, C. S. & Begelman, M. C. X-ray signatures of evolving radio galaxies. *Astrophys. J.* **501**, 126–136 (1998).
9. Kaiser, C. R. & Alexander, P. Heating of the intergalactic medium by FR II radio sources. *Mon. Not. R. Astron. Soc.* **305**, 707–723 (1999).
10. Reynolds, C. S., Heinz, S. & Begelman, M. C. The hydrodynamics of dead radio galaxies. *Mon. Not. R. Astron. Soc.* (in the press).
11. Fabian, A. C. *et al.* Chandra imaging of the complex X-ray core of the Perseus cluster. *Mon. Not. R. Astron. Soc.* **318**, L65–L68 (2000).
12. Brüggemann, M. & Kaiser, C. R. Buoyant radio plasma in clusters of galaxies. *Mon. Not. R. Astron. Soc.* **325**, 676–684 (2001).
13. McNamara, B. R. *et al.* Discovery of ghost cavities in Abell 2597's X-ray atmosphere. *Astrophys. J.* (submitted); preprint astro-ph/0110554 at (<http://xxx.lanl.gov>) (2001).
14. Brüggemann, M., Kaiser, C. R., Churazov, E. & Ensslin, T. A. Simulation of radio plasma in clusters of galaxies. *Mon. Not. R. Astron. Soc.* (in the press).
15. Quilis, V., Bower, R. G. & Balogh, M. Bubbles, feedback and the intra-cluster medium: Three-dimensional hydrodynamic simulations. *Mon. Not. R. Astron. Soc.* **328**, 1091–1097 (2001).
16. Fryxell, B. *et al.* FLASH: An adaptive mesh hydrodynamics code for modeling thermonuclear flashes. *Astrophys. J. Suppl.* **131**, 273–334 (2000).
17. Nulsen, P. E. J. & Böhringer, H. A ROSAT determination of the central mass of the Virgo cluster. *Mon. Not. R. Astron. Soc.* **274**, 1093–1106 (1995).
18. Binney, J. & Tabor, G. Evolving cooling flows. *Mon. Not. R. Astron. Soc.* **276**, 663–678 (1995).
19. Soker, N., White, R. E., David, L. P. & McNamara, B. R. A moderate cluster cooling flow model. *Astrophys. J.* **549**, 832–839 (2001).

Acknowledgements

The computations reported here were performed using the UK Astrophysical Fluids Facility (UKAFF). The software used in this work was in part developed by the DOE supported ASCI/Alliances Center for Thermonuclear Flashes at the University of Chicago.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.B. (e-mail: m.brueggemann@iu-bremen.de).

Plasmon-assisted transmission of entangled photons

E. Altewischer, M. P. van Exter & J. P. Woerdman

Leiden University, Huygens Laboratory, PO Box 9504, 2300 RA Leiden, The Netherlands

The state of a two-particle system is said to be entangled when its quantum-mechanical wavefunction cannot be factorized into two single-particle wavefunctions. This leads to one of the strongest counter-intuitive features of quantum mechanics, namely non-locality^{1,2}. Experimental realization of quantum entanglement is relatively easy for photons; a starting photon can spontaneously split into a pair of entangled photons inside a nonlinear crystal. Here we investigate the effects of nanostructured metal optical elements³ on the properties of entangled photons. To this end, we place optically thick metal films perforated with a periodic array of subwavelength holes in the paths of the two entangled photons. Such arrays convert photons into surface-plasmon waves—optically excited compressive charge density waves—which tunnel through the holes before reradiating as photons at the far side^{4–7}. We address the question of whether the entanglement survives such a conversion process. Our coincidence counting measurements show that it does, so demonstrating that the surface plasmons have a true quantum nature. Focusing one of the photon beams on its array reduces the quality of the entanglement. The propagation of the surface plasmons makes the array effectively act as a ‘which way’ detector.

Light can potentially couple to surface plasmons (SPs) if the surface in which the SPs reside shows a periodic structure to satisfy conservation of energy and momentum. The samples that we have studied are two metal hole arrays. Each array is 1 mm × 1 mm, and comprises a 200-nm-thick gold layer (on glass) perforated with a square grid of 200-nm-diameter holes spaced with a 700-nm lattice constant; Fig. 1a inset shows a typical scanning electron microscope picture. This figure also shows transmission spectra of the two arrays, measured with a spectrometer using normally incident white light. We can clearly see the resonances due to excitation of SPs on either of the metal–dielectric boundaries. At these resonances, the measured transmission can be orders of magnitude larger than the value obtained from classical diffraction theory for subwavelength holes^{4,8}. In a simple picture, the surprisingly large transmission is due to the coupling of a photon to an SP on one side of the metal, subsequent tunnelling of the SP through the holes to establish an SP at the other side, and final reradiation into a photon⁵. Other prominent features in the spectra are the transmission minima associated with Wood anomalies⁶. The theoretical description of the full transmission spectrum is incomplete at present, but the role of the SP is well established^{5–7}. The resonance used in our experiment is centred at 809 nm, and has a width of ~40 nm; a calculation based on the geometry of the array and the optical constants of gold and glass shows that it is associated with the $(\pm 1, \pm 1)$ SP mode on the glass–metal interface. The label $(\pm 1, \pm 1)$ denotes any of the four diagonal directions that are frequency degenerate for excitation under normal incidence. Peak transmissions of the two arrays at 813 nm are typically 3% (dashed curve) and 5% (solid curve); these values are much larger than the value of 0.55% obtained from classical theory⁸. The difference in transmission between the two nominally identical hole arrays is ascribed to production imperfections.

Because an SP is a mainly longitudinal, compressive electron density wave, its propagation direction depends on the polarization axis of the incident light, following a certain dispersion relation^{9,10}.

In order to confirm this for our samples, we have measured polarization-resolved transmission spectra of one of the hole arrays for various angles of incidence θ_{inc} of plane-wave radiation (Fig. 1b and c). Angle tuning is expected to shift the various resonances in the transmission spectrum in different ways. As we have theoretically associated the peak at 809 nm with a $(\pm 1, \pm 1)$ SP, we have varied the angle of incidence by tilting the samples along the diagonal axis of the square hole array. For incident light polarized orthogonal to this tilting axis (Fig. 1b), the 809-nm peak splits and shifts for increasing θ_{inc} ; for incident light polarized along the tilting axis (Fig. 1c), the 809-nm peak remains at approximately the same spectral position. These results confirm the association of the 809-nm peak with the $(\pm 1, \pm 1)$ SP.

We generate polarization-entangled twin photons at a wavelength of 813 nm with the standard method of type II spontaneous parametric down-conversion^{11,12} depicted in Fig. 2. The twin

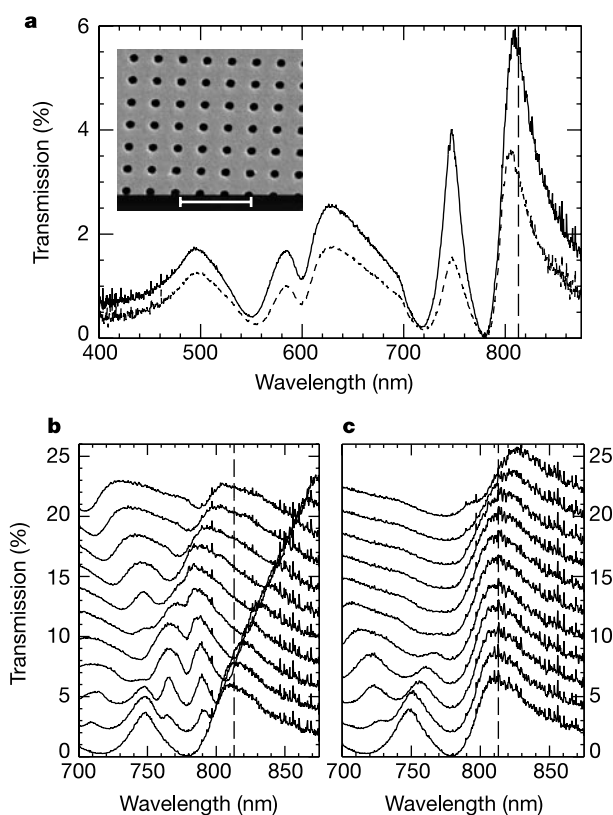


Figure 1 Wavelength-dependent transmission of the arrays used in the experiment. The dashed vertical line indicates the wavelength of 813 nm used in the entanglement experiment. **a**, The transmission of the two arrays. Inset, scanning electron microscope picture of part of a typical hole array. Scale bar, 2 μm. The hole arrays were produced at the Delft Institute of Micro-Electronics and Sub-micron Technology (DIMES) by first defining, with electron beam lithography, arrays of dielectric pillars on a 0.5-mm-thick glass substrate, subsequently evaporating the gold layer onto a 2-nm titanium bonding layer, and finally etching away the pillars to leave the array of air holes. **b**, **c**, Wavelength-dependent transmission of one of the hole arrays for various angles of incidence at a polarization orthogonal to the tilting axis (**b**) and along the tilting axis (**c**). In both graphs the lowest curve is measured with the probe beam at normal incidence, while consecutive curves are measured at increasing angles of incidence (one-degree steps) by subsequent tilting of the square hole array around a diagonal. These curves have been plotted with subsequent vertical offsets of 2% on the transmission scale. The resonance around 813 nm (dashed line) shows a complicated splitting for a polarization orthogonal to the tilting axis (**b**), whereas it is approximately stationary for a polarization along this axis (**c**).

photons travel along two ‘beam lines’ defined by pinholes, pass through polarizers P1 and P2, and are detected by single-photon counters. The photodetectors act as ‘bucket detectors’—that is, they impose no further transverse mode selection (this is only done by pinholes). To provide the two-photon coincidence rate, the signals from the two detectors are combined electronically in a coincidence circuit with a time window of 2 ns.

In a simplified picture, the generated polarization-entangled state is

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|H_1V_2\rangle + e^{i\theta}|V_1H_2\rangle) \quad (1)$$

where the state $|H_1V_2\rangle$ represents the simultaneous emission and propagation of an H-polarized photon in beam 1 and a V-polarized photon in beam 2. The H- and V-directions are defined by the orthogonal birefringent axes of the BBO (β barium borate) crystal generating the twin photons, and all spatial information is implicitly contained in the beam labelling. By tilting one of the compensating crystals (C in Fig. 2), the quantum phase θ can be set.

In the absence of the hole arrays, our set-up produces typically 3.2×10^4 coincidence counts per second, which is $\sim 25\%$ of the single count rate. A measure for the purity of the quantum entangled state is the so-called visibility of the biphoton fringe^{11,12}. This visibility was typically $V_{0^\circ} = 99.3\%$ and $V_{45^\circ} = 97.0\%$ at polarization angles of 0° and 45° , respectively (see Table 1). The high value at 45° shows that the natural preference of the BBO crystal for its birefringent axes (0° and 90°) was almost completely

removed in our set-up by the compensating crystals C, which act as quantum erasers.

Placement of the two hole arrays in the two beam lines leads naturally to a marked reduction of single and coincidence counts. Coincidence count rates are typically 55 s^{-1} at the optimum setting of the detecting polarizers, which is consistent with the transmissions of the arrays given above ($3\% \times 5\% \times 3.2 \times 10^4 \text{ s}^{-1} \approx 50 \text{ s}^{-1}$). We again measured the purity of the entangled state, and found that the visibilities were now $V_{0^\circ} = 97.1\%$ and $V_{45^\circ} = 97.2\%$, respectively. In Fig. 3 the corresponding fourth-order quantum interference fringes are shown for polarizer P2 fixed at 0° and 45° , and P1 varying in steps of 10° . These measurements show convincingly that the quantum entanglement almost completely survives the transition from photon to SP and back. As a further confirmation, we performed a measurement of the so-called S-parameter, as described in ref. 11, on a singlet Bell state ($\theta = \pi$ in equation (1)). This experiment, which took 16 runs of 100 seconds each, gave a value of $|S| = 2.71 \pm 0.02$, which is a violation of the classical inequality ($|S| \leq 2$; ref. 13) by 35 standard deviations.

Further experiments were done on a set-up with only a single array in one of the beams. The results for this case look very similar to those for the experiment with two arrays; again the entanglement was practically unaffected (see Table 1). This is to be expected, as the two-photon wavefunction of equation (1) is perturbed by changes in either of its single-photon components; in principle, a single array could have removed all entanglement. The difference between the two single-array experiments (Table 1) is due to imperfections in array 2, which are also observable in its (single-photon) polarization-dependent transmission. As the measurements using only array 2 gave results very similar to the situation with both arrays in place, these imperfections must have caused the somewhat limited visibilities in the two-array experiment.

The most intriguing results of our single-array experiment are obtained when we focus one of the beam lines onto its hole array, using a confocal telescope (close to lens L) of two $f = 15 \text{ mm}$ lenses symmetrically positioned around the array, as shown inside the large dotted rectangle in Fig. 2. Under these conditions, we observe a notable reduction of the degree of quantum entanglement: when the intra-telescope focus has a numerical aperture of 0.13, we observe visibilities of $V_{0^\circ} = 73\%$ and $V_{45^\circ} = 87\%$ (Table 1).

The observed reduction in visibility on focusing can be explained as a consequence of the non-local relation between the electronic excitation in the metal film and the incident optical field; SPs are not at all local, but instead propagate along the dielectric–metal interface at nearly the speed of light over distances of many optical wavelengths^{3,9,10}. As a result of this propagation, the near-field distribution of the photons that are reradiated at the back of the hole array differs from the spatial profile of the ‘polarization-isotropic’ incident photons. Because we use the $(\pm 1, \pm 1)$ SP resonance (at 809 nm) we expect—for unpolarized incident light—a near-field pattern consisting of two orthogonal ‘ellipses’ at the back of the hole array (Fig. 2 inset). The unpolarized overlap region of these ellipses correspond to the focused incident light, whereas the polarized ‘protuberances’ introduce the possibility of distinguishing the polarization of the photons on the basis of their

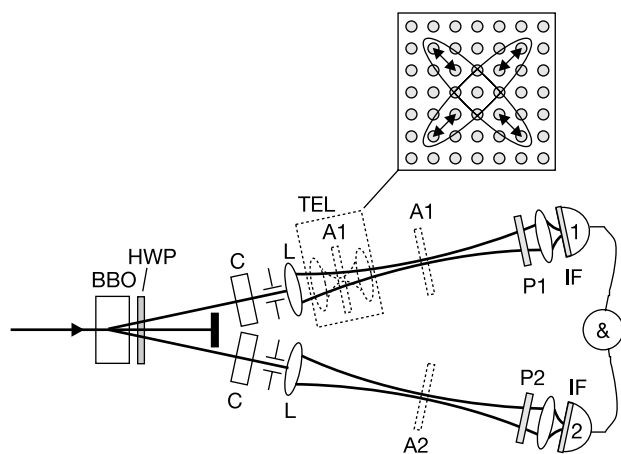


Figure 2 Experimental set-up. A 240-mW continuous-wave krypton-ion laser beam at a wavelength of 406.7 nm is directed onto a 1-mm-thick BBO nonlinear crystal, where the beam diameter is $\sim 0.50 \text{ mm}$ (full width at $1/e^2$ points). Inside the nonlinear crystal, a small fraction of the pump photons is down-converted into twin photons at the doubled wavelength (813 nm); these are emitted along two intersecting cones. Polarization-entangled photon pairs are selected with pinholes at the crossings of these cones; the size of the pinholes (far-field diameter 5 mrad) was chosen as a compromise between high yield and good quantum entanglement. Lenses L of 40 cm focal length produce a one-to-one intermediate image of the pumped area, which is used in some experiments to accommodate the hole arrays A1 and A2. After passing through polarizers P1 and P2, the entangled twin photons are focused through interference filters IF (10-nm bandwidth centred at 813 nm) onto single-photon counting modules (Perkin Elmer SPCM-AQR-14). Beam walkoff is compensated by the standard quantum eraser comprising a half-wave plate HWP at 45° and compensating crystals C with a thickness equal to half of that of the generating crystal^{11,12}. The dotted objects are present only in some experiments; they show the hole arrays A1 and A2 at the image positions created by lenses L, or, alternatively, in the focus of the confocal telescope TEL (15 mm focal length lenses). Inset, schematic picture of the near field at the back of array A1 when this is positioned inside the telescope. The arrows indicate the polarization direction of the optical electric field; the centre region is unpolarized.

Table 1 Biphoton fringe visibilities

Experiment	$R \text{ (s}^{-1}\text{)}$	$V_{0^\circ} \text{ (%)}$	$V_{45^\circ} \text{ (%)}$
No arrays	32×10^3	99.3	97.0
Both arrays	55	97.1	97.2
Only array 1	1.6×10^3	99.4	97.1
Only array 2	1.0×10^3	97.5	96.8
Array 1, focussed	1.1×10^3	73	87

R , measured coincidence count rate; V_{0° and V_{45° , measured visibility for one of the polarizers fixed at 0° and 45° , respectively.

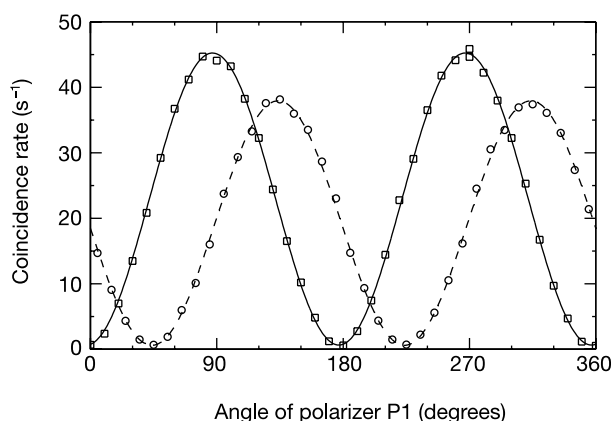


Figure 3 Biphoton fringes. These fringes correspond to fourth-order quantum interference, and were measured with the two hole arrays in place, for P2 fixed at 0° (solid curve) and 45° (dashed curve), and P1 varying in steps of 10°.

spatial near-field coordinates. This will then automatically remove part of the polarization entanglement, just as any ‘which way’ information will do. Note that the observed reduction in visibility is much stronger for V_{0° than for V_{45° , contrary to what is generally found without using hole arrays^{11,12}. This observation is consistent with the fact that we excite SPs propagating in the ‘diagonal’ (45°) directions.

The non-local nature of the electronic response is equivalent to an explicit wavevector dependence of the dielectric function (‘spatial dispersion’; ref. 14). The latter description highlights the far-field aspects of a non-local dielectric response, and is related by way of a Fourier transform to the near-field picture given above.

A theory for the reduction of visibility due to this effect must intrinsically be a multimode theory because, after passage through the hole arrays, the spatial information within each beam should be accounted for. Discussion of such a theory is beyond the scope of this Letter, but qualitative arguments are easily found. When adopting a near-field point of view, the reduction of the visibility is expected to be significant when the propagation or coherence length l of the SP is comparable to the transverse coherence length of the focused incident beam. This is indeed the case for our experiment: on the basis of the spectral width of the transmission peak at 809 nm, we estimate $l \approx 4 \mu\text{m}$; the number of holes ‘covered’ by the propagating SPs is thus about 10. In comparison, the transverse coherence length of the light is $\sim 4 \mu\text{m}$ (set by the 5-mrad far-field selection due to the pinholes; Fig 2). We note that the SP propagation length is an order of magnitude smaller than the value $l \approx 40 \mu\text{m}$ predicted for a smooth pure gold film³; we ascribe this difference to radiative decay due to the hole patterning, and to extra damping by the titanium bonding layer.

From a general perspective, the observed conservation of quantum entanglement for the conversion from photon $\rightarrow$ surface plasmon $\rightarrow$ photon is a demonstration of the true quantum nature of SPs. All experiments on SPs so far have not probed their quantum nature, but only their wave nature through their semiclassical dispersion relation. We note that the true quantum nature of the photon was only established in 1977, in anti-bunching experiments^{15,16}. Furthermore, a simple estimate shows that SPs are very macroscopic, in the sense that they involve some 10^{10} electrons. Our experiment proves that this macroscopic nature does not impede the quantum behaviour of SPs, because they can act as intermediates in transmitting entangled photons to yield the expected fourth-order interference.

A theory for our experiments based only on locally induced surface currents is clearly inadequate. We stress this point because some recent theoretical models for one-dimensional gratings in thick metal films have questioned the role of SPs, emphasizing

instead waveguide effects. Our arrays are apparently thin enough (thickness/period ≈ 0.3) for waveguide effects not to play an important role. This conclusion is supported by experiments in which the thickness of such a ‘thin’ array has been varied⁶.

By addressing the topic of plasmon-assisted transmission of quantum entangled photons, we have combined two fields of research, namely quantum information and nanostructured metal optics. We hope that our work will stimulate other studies of the transfer of entanglement to condensed-matter degrees of freedom. $\square$

Received 11 March; accepted 23 May 2002; doi:10.1038/nature00869.

- Greenberger, D. M., Horne, M. A. & Zeilinger, A. Multiparticle interferometry and the superposition principle. *Phys. Today*, 22–29 (August 1993).
- Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
- Raether, H. *Surface Plasmons* (Springer, Berlin, 1988).
- Ebbesen, T. W., Lezec, H. J., Ghaemi, H. F., Thio, T. & Wolff, P. A. Extraordinary optical transmission through sub-wavelength hole arrays. *Nature* **391**, 667–669 (1998).
- Martin-Moreno, L. *et al.* Theory of extraordinary optical transmission through subwavelength hole arrays. *Phys. Rev. Lett.* **86**, 1114–1117 (2001).
- Ghaemi, H. F., Thio, T., Grupp, D. E., Ebbesen, T. W. & Lezec, H. J. Surface plasmons enhance optical transmission through subwavelength holes. *Phys. Rev. B* **58**, 6779–6782 (1998).
- Grupp, D. E., Lezec, H. J., Ebbesen, T. W., Pellerin, K. M. & Thio, T. Crucial role of metal surface in enhanced transmission through subwavelength apertures. *Appl. Phys. Lett.* **77**, 1569–1571 (2000).
- Bethe, H. A. Theory of diffraction by small holes. *Phys. Rev.* **66**, 163–182 (1944).
- Hecht, B., Bielefeldt, H., Novotny, L., Inoué, Y. & Pohl, D. W. Local excitation, scattering, and interference of surface plasmons. *Phys. Rev. Lett.* **77**, 1889–1892 (1996).
- Sönnichsen, C. *et al.* Launching surface plasmons into nanoholes in metal films. *Appl. Phys. Lett.* **77**, 140–142 (2000).
- Kwiat, P. G. *et al.* New high-intensity source of polarization-entangled photon pairs. *Phys. Rev. Lett.* **75**, 4337–4341 (1995).
- Kurtsiefer, C., Oberparleiter, M. & Weinfurter, H. High efficiency entangled photon pair collection in type II parametric fluorescence. *Phys. Rev. A* **64**, 023802–1–023802–4 (2001).
- Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
- Landau, L. D., Lifshitz, E. M. & Pitaevskii, L. P. *Electrodynamics of Continuous Media* 2nd edn (Pergamon, Oxford, 1984).
- Lu, Y. J. & Ou, Z. Y. Observation of nonclassical photon statistics due to quantum interference. *Phys. Rev. Lett.* **88**, 023601–1–023601–4 (2002).
- Kimble, H. J., Dagenais, M. & Mandel, L. Photon antibunching in resonance fluorescence. *Phys. Rev. Lett.* **39**, 691–695 (1977).

Acknowledgements

We thank A. van Zuuk and E. van der Drift for the production of the hole arrays, and G. Nienhuis for theoretical discussions. This work was supported by the Stichting voor Fundamenteel Onderzoek der Materie (FOM), and the European Union under the IST-ATESIT contract.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to E.A. (e-mail: erwin@molphys.leidenuniv.nl).

Atomistic mechanisms governing elastic limit and incipient plasticity in crystals

Ju Li^{*†}, Krystyn J. Van Vliet[‡], Ting Zhu^{*}, Sidney Yip^{*‡} & Subra Suresh[‡]

^{*} Department of Nuclear Engineering, [‡] Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

[†] These authors contributed equally to this work

Nanometre-scale contact experiments^{1–6} and simulations^{7–10} demonstrate the potential to probe incipient plasticity—the onset of permanent deformation—in crystals. Such studies also point to the need for an understanding of the mechanisms governing defect nucleation in a broad range of fields and applications. Here we present a fundamental framework for describing incipient plasticity that combines results of atomistic and finite-element modelling, theoretical concepts of structural stability at finite strain, and experimental analysis. We quantify two key features of the nucleation and subsequent evolution of defects. A position-sensitive criterion based on elastic stability determines the location and character of homogeneously nucleated defects. We validate this stability criterion at both the atomistic and the continuum levels. We then propose a detailed interpretation of the experimentally observed sequence of displacement bursts to elucidate the role of secondary defect sources operating locally at stress levels considerably smaller than the ideal strength required for homogeneous nucleation. These findings provide a self-consistent explanation of the discontinuous elastic–plastic response in nanoindentation measurements, and a guide to fundamental studies across many disciplines that seek to quantify and predict the initiation and early stages of plasticity.

Characteristic discontinuities have been observed consistently in the measured load–penetration depth (P – h) response of single crystals indented to nanometre-scale depths^{1–5}. Before the first discontinuity, local shear stresses beneath the nominally sharp indenter approach the theoretical strength of the indented crystal, indicating that homogeneous defect nucleation is a logical starting event for subsequent incipient (that is, early-stage) plasticity (Fig. 1a). This hypothesis is supported by recent *in situ* experiments⁶ using the Bragg–Nye bubble raft analogue, showing that nanoindentation of a two-dimensional (2D) crystal indeed results in homogeneous dislocation nucleation. However, the present level of quantitative understanding of the mechanisms by which contact-induced plasticity initiates and evolves in three-dimensional (3D) crystals is still limited.

The displacement bursts in the load-controlled experiment of Fig. 1a would appear as sharp decreases in load (dips) in a displacement-controlled experiment, as shown for a simulated displacement-controlled response in Fig. 1b. These dips correspond to elastic instabilities that can be predicted by a free-energy-based, position-dependent criterion, and its validity demonstrated at both the atomic and continuum levels. To show this, we consider a representative volume element V subjected to homogeneous deformation at finite strain to a current configuration $\mathbf{x}$. Expanding the free energy F to the second order in incremental displacement $\mathbf{u}(\mathbf{x})$, we obtain (J.L. *et al.*, manuscript in preparation)

$$\Delta F = \frac{1}{2} \int_{V(\mathbf{x})} D_{ijkl} u_{i,j}(\mathbf{x}) u_{k,l}(\mathbf{x}) dV \quad (1)$$

where $D_{ijkl} \equiv C_{ijkl} + \tau_{ji} \delta_{ik}$, C is the elastic constants tensor, τ is the

internal or Cauchy stress tensor, and $u_{i,j} \equiv \partial u_i(\mathbf{x}) / \partial x_j$. Substitution of a plane wave perturbation, $u_i(\mathbf{x}) = w_i \exp(i\mathbf{k} \cdot \mathbf{x})$, into equation (1) leads to the following stability condition for the representative volume element:

$$\Lambda(\mathbf{w}, \mathbf{k}) \equiv (C_{ijkl} w_i w_k + \tau_{ji}) k_j k_l > 0 \quad (2)$$

The sign of $\Lambda(\mathbf{w}, \mathbf{k})$ reflects the concavity of F ($\mathbf{w}$ is the polarization vector, and $\mathbf{k}$ the wavevector). If there exists a pair of $\mathbf{w}, \mathbf{k}$ such that $\Lambda(\mathbf{w}, \mathbf{k})$ is negative, then homogeneity of this representative volume element cannot be maintained and defect singularities will form internally in four stages. (1) Linear growth of the unstable elastic wave, well-described by continuum; (2) non-linear evolution at larger amplitudes, over which the wave profile steepens; (3) progressive steepening of the wave front until its width approaches atomic spacing, upon which its description must be transferred from a continuum to an atomistic basis; and (4) arrest of the atomistically sharp wave front in a low-dimensional atomic

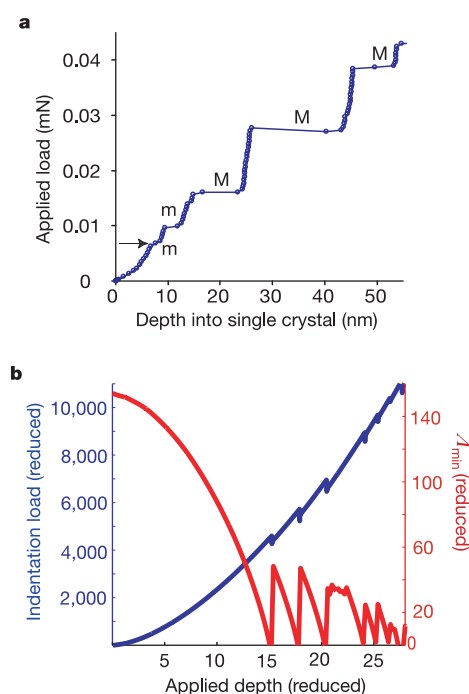


Figure 1 Nanoindentation responses. **a**, Experimentally measured applied load–displacement response of {111} single-crystal aluminium⁴ showing an initial displacement burst (arrow). Following this first event, multiple displacement bursts can be observed. The displacement bursts are either minor (m), spanning ~2 nm or less, or major (M), 10 nm or more, both at essentially constant applied load. A minor burst indicates a single slip event whereby just one prismatic or glide loop is created, which then propagates until being stopped at the substrate (or at a grain boundary in the case of a polycrystal). In contrast, a major burst indicates at least ten consecutive complete slips, at no observable increase in applied load. Just one such major burst would exhaust at least a significant portion of the elastic energy stored in a particular slip system⁴. **b**, Applied displacement–load response (blue) of MD simulated bubble raft, showing correlation of load drops with a sequence of observed homogeneous dislocation dipole nucleation events. The elastic instability criterion $\Delta_{\min}$ (red) tends to zero for the first several nucleation events. All nucleation events occur subsurface; the high mobility of the dissociated dipoles immediately restores the region beneath the indenter to crystalline order after nucleation. Our interbubble potential is of the form $V(r) = \alpha(r-1)^4 - \beta(r-1)^2$, where r is the interbubble distance. Length and energy units represented by the cut-off radius and potential well-depth, respectively, are set equal to unity such that the potential well shape is controlled by a single parameter, the equilibrium distance r_0 , set equal to 0.85. The indenter of radius $R = 160$ is implemented as a frictionless, external potential on the bubbles. The bubble raft is 340×212 in dimension, containing 115,200 bubbles for this in-plane orientation.

energy profile, at which point a defect is nucleated. If the wave is transverse, the instability will probably result in the formation of a dislocation or twin of slip plane normal $\mathbf{k}$ and Burgers vector orientation $\mathbf{w}$; if the wave is longitudinal, a microcrack would result. The elastic stability of the representative volume element thus can be determined by minimizing $\Lambda(\mathbf{w}, \mathbf{k})$ with respect to the polarization vector $\mathbf{w}$ and wavevector $\mathbf{k}$, $\Lambda_{\min} \equiv \min_{|\mathbf{w}|=1, |\mathbf{k}|=1} \Lambda(\mathbf{w}, \mathbf{k})$. By interpreting $\Lambda_{\min}$ as a measure of micro-stiffness that varies locally, instability is predicted at the position where $\Lambda_{\min}$ vanishes. Note that equation (2) is an energy-based criterion; its minimization at a material point is, in fact, environment-dependent. The present approach is different from the analysis of instability under homogeneous loading^{11–13} in two respects. In the case of nanoindentation, stress applied via the frictionless indenter is nonuniform and thus can create an elastically unstable volume that is isolated physically from the external loading. Secondly, our derivation is the first (to our knowledge) such expression capable of predicting not only the location of elastic instability, but also the slip character (slip plane and Burgers vector) of the corresponding defect, as shown below.

As a criterion derived from the free energy, equation (2) is equivalent to the continuum-based condition discussed formally by Hill¹⁴ in the context of acceleration waves and by Rice¹⁵ when considering shear band formation. In our implementation we ascribe to each atom an atomistic stress tensor¹⁶ and an elastic constant tensor¹⁷, and identify the ‘softest atom’ by minimizing $\Lambda(\mathbf{w}, \mathbf{k})$ for each atom iteratively, first with respect to $\mathbf{w}$ and then

with respect to $\mathbf{k}$ to obtain a self-consistent $\Lambda_{\min}$. The lowest $\Lambda_{\min}$ found at a given indenter displacement is shown for a 2D, molecular dynamics (MD) simulation of the Bragg–Nye raft (Fig. 1b). During elastic deformation, $\Lambda_{\min}$ decreases monotonically and vanishes exactly when the first load dip occurs. The correspondence between a load dip and the vanishing of $\Lambda_{\min}$ continues for five subsequent dips. Furthermore, we find that the $\mathbf{w}$ and $\mathbf{k}$ vectors that minimize $\Lambda(\mathbf{w}, \mathbf{k})$ agree well with the slip direction and slip plane normal, respectively, observed in MD simulations, which are also consistent with the primary slip system of the face-centred cubic crystallographic lattice. Considering the complexity of the numerical evaluations, we regard the results of Fig. 1b as significant and quantitative validation of the stability criterion in both formulation and implementation. To our knowledge, this is the first demonstration of a site-specific stability criterion at the atomic scale.

We have extended this stability analysis to finite element calculations based on Cauchy–Born elasticity¹⁸ by incorporating large strain constitutive relations provided directly by interatomic potentials. In contrast to the quasicontinuum method developed by Tadmor *et al.*¹⁹, which in principle should give the same result for the present nucleation problem, our approach remains a fully continuum method that is capable of describing the elastic instabilities predicted by the Λ -criterion. Its implementation in a standard finite element software package²⁰ requires only that the user supply the exact Cauchy–Born constitutive relation, which comes directly from a static (temperature $T = 0$) or ensemble-averaged lattice sum calculating at every step with the appropriate interatomic potential. This method allows us to identify the lowest $\Lambda_{\min}$ among all nodes,

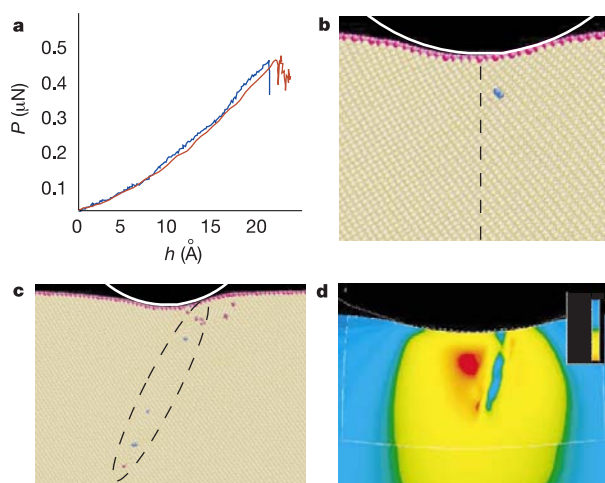


Figure 2 Correspondence of MD and FEM simulations in $\langle 111 \rangle$ cylindrical indentation of Cu. In both simulations, the long axis of the cylinder is along $\langle 110 \rangle$. **a**, Applied displacement–load responses (h – P) showing correlation of load relaxation with homogeneous defect nucleation. The FEM response (blue) shows good agreement with MD (red) for the initial, elastic loading response and the critical load required to nucleate a defect. **b**, MD simulation comprising 290,304 atoms and periodic boundary conditions, showing that the nucleation site (blue) is off the centre axis (dashed line) by $\sim 20\%$, and at a depth $\sim 60\%$ of the contact radius a . Indenter radius $R = 20$ nm. Atoms are colour-encoded by coordination number N : yellow, $N = 12$; blue, $N \neq 12$. The lack of left–right symmetry and decreased depth of nucleation are due to the elastic anisotropy of Cu, and are consistent with the recently developed linear elastic, anisotropic solution for a cylindrical punch²³. **c**, MD simulation after several dipole nucleation events illustrating the formation of a shear band (dashed lines) after the dissociation of dislocation pairs to the free surface and crystalline interior. **d**, FEM simulation of the initial defect nucleation event (elements are colour-encoded by the Mises stress, red and blue denote maximum and minimum values respectively) showing that shear localization spontaneously appears immediately after $\Lambda_{\min} = 0$. The predicted nucleation site, slip plane normal and Burgers vector all agree with MD observations after the event.

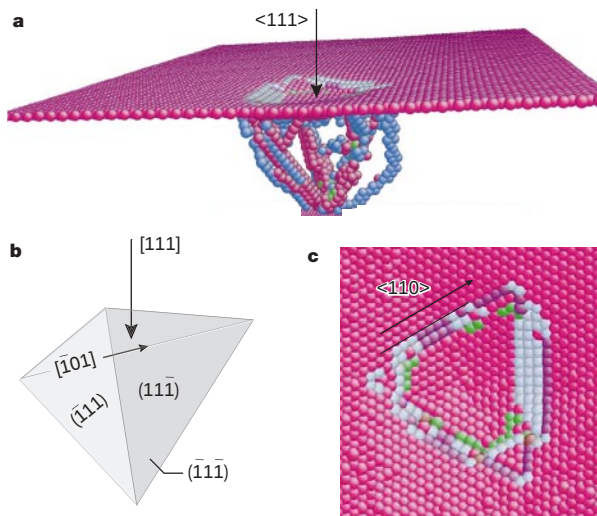


Figure 3 Homogeneous defect nucleation in $\langle 111 \rangle$ spherical indentation of Al. **a**, MD simulation of spherical indentation ($R = 12$ nm) normal to the $\langle 111 \rangle$ surface of Al results in a sessile structure derived from homogeneously nucleated glide loops on three equivalent $\{111\}$ planes. The simulation comprises 326,592 atoms and periodic boundary conditions, using the Ercolessi potential²⁴ for Al. Atom colour indicates coordination number N : light pink, surface coordination, $N = 9$; atoms with perfect coordination $N = 12$ are made invisible. Other colours (green, blue and light pink interior atoms) indicate under- or over-coordinated atoms, which are atoms comprising defects. **b**, Diagram of the $\{111\}$ slip systems available for indentation in the $\langle 111 \rangle$ direction. Symmetric arrangement of the systems indicates that homogeneous nucleation of dislocation loops on multiple slip systems must result in loop intersection and, consequently, sessile dislocation segments in the interior of the crystal. **c**, Top view of the MD simulation surface showing the $\langle 110 \rangle$ orientations of three slip steps (surface ledges) that result from the initial nucleation event. The surface ledge corners then act as stress concentrators for further heterogeneous nucleation events.

thereby predicting homogeneous defect nucleation in accordance with an atomic-scale strain-to-failure criterion. Direct comparison of the load–displacement responses for cylindrical $\langle 111 \rangle$ indentation of Cu obtained by direct MD and the aforementioned finite element method (FEM) using the same interatomic potential²¹ is shown in Fig. 2a. The close correspondence in the elastic loading and the critical load for defect nucleation given by these two traditionally different methods of modelling is noteworthy. Further validation of our interatomic-potential-based FEM calculation against direct MD under the condition of large nonlinear strain and homogeneous defect nucleation is demonstrated by the fidelity of the nucleation site, slip plane, and Burgers vector obtained from these two approaches (Fig. 2b–d).

We next consider MD simulations of spherical indentation, showing the atomic-level details of dislocation nucleation and subsequent incipient plasticity. Figure 3a shows that defect nucleation during displacement-controlled $\langle 111 \rangle$ nanoindentation of Al occurs near the surface as expanding glide loops on three equivalent $\{111\}\langle 110 \rangle$ slip systems (Fig. 3b). Kelchner *et al.* reported initial glide loop nucleation for atomistic simulations of spherical $\langle 111 \rangle$ indentation in another face-centred cubic crystal (Au), but did not observe the three-fold symmetry of the defect structure⁷. This contrast may be due to differences in elastic anisotropy of the crystal, stacking fault energy, and the computational method employed (Kelchner *et al.* used energy minimization, as opposed to MD). Owing to the inward orientation of the $\{111\}$ planes activated during indentation of Al, the glide loops expand and intersect to produce three, intersecting $\langle 110 \rangle$ -oriented slip steps on the surface (Fig. 3c). Upon increasing indenter displacement, loops nucleate on adjacent $\{111\}$ planes; four such loops interact, with a segment from each forming one side of a parallelogram (Fig. 4a). This prismatic loop structure moves through the crystal along its glide prism to the bottom of the substrate (Fig. 4b and Supplementary Information Fig. 1). It is important to note that the prismatic loops are not homogeneously nucleated; instead they are emitted near the surface, aided by the reaction products of homogeneously nucleated glide loops. As shown in 3D simulations for Cu (Supplementary Information Fig. 2), increased elastic anisotropy alters the lock/source formation in $\langle 111 \rangle$ indentation, leading to an

incipient plasticity mode more similar to 2D observations, that is, expansion of glide loops and twinning.

We now analyse the experimental observations of $\{111\}$ single-crystal Al (ref. 4), indented with a nominally sharp indenter (approximate indenter tip radius $R = 50$ nm), as shown in Fig. 1a. The P – h response is elastic until a shear stress of ~ 5.5 GPa is attained beneath the rounded tip. If both the initial and subsequent slip events were homogeneously nucleated, then each subsequent event would require gigapascal-level local stresses observed as an increase in applied load (hardening). Experimentally, however, no apparent hardening is observed during major displacement bursts, which correspond to the creation and propagation of ~ 50 complete slip events⁴. Thus, a hypothesis of multiple homogeneously nucleated prismatic loops is incompatible with the experimental data on major bursts. With reference to our MD simulations, we suggest that incipient plasticity is initiated by homogeneously nucleated glide loops that subsequently react to form 3D, secondary dislocation sources similar to the Frank–Read source. These sources operate at megapascal-level local stresses until most of the stored elastic energy is exhausted. After significant emission from a single secondary source, back-stresses from dislocation pile-up counteract the external resolved stress for the corresponding slip system, thereby shutting down this dislocation source. Subsequent bursts occur only when the applied load is increased sufficiently to cause existing sources to emit single dislocation loops discretely, resulting in minor relaxations. The minor bursts and hardening continue until the creation of a source on a different slip system, leading to a second major relaxation, and this process repeats. This interpretation, supported by recent atomistic calculations which show that inhomogeneities such as atomic steps and grooves near the indented surface can greatly reduce the nucleation stress in nanoindentation^{8,22}, provides a self-consistent, mechanism-based explanation of the observed transition from the onset of elastic instability to incipient plasticity in nanoindentation responses such as Fig. 1a.

Although our study is framed in the context of indentation, the quantitative and predictive capabilities of our results have implications for both experimental and computational investigations of large-strain deformation phenomena. For example, dislocation

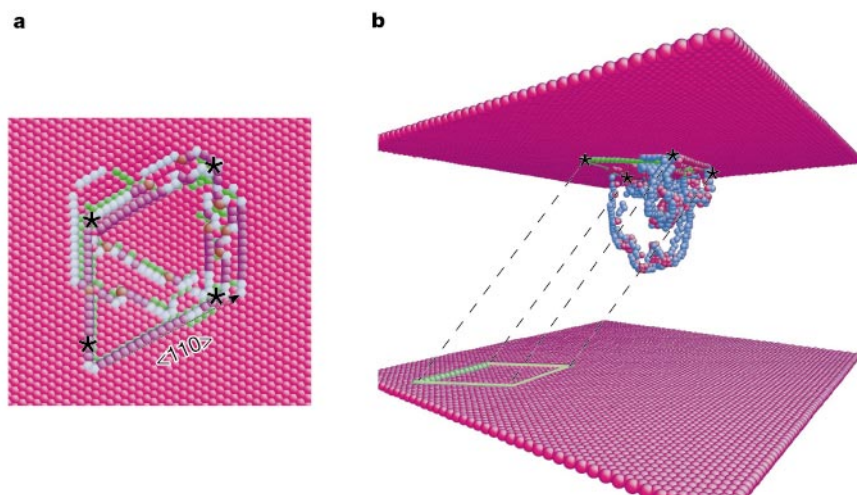


Figure 4 Incipient plasticity in $\langle 111 \rangle$ spherical indentation of Al. **a**, Plan view of the MD simulation comprising 350,000 atoms. The intersection points (stars) indicate the corners of the emitted prismatic loop. **b**, On further indenter displacement, dislocation half-loops on two $\{111\}\langle 110 \rangle$ slip systems cross-slip and pinch off from the surface, forming a prismatic dislocation loop which moves through the crystal along its glide prism

(dashed lines) without increasing in loop diameter. Atom colour corresponds to coordination number N as described in Fig. 3 legend. A parallelogram slip mark is left on the bottom surface when the prismatic loop leaves the lower, unconstrained film surface.

dynamics studies of strain hardening that at present assume an initial dislocation configuration could incorporate the proposed interatomic-potential-based FEM to solve for the regular stress field solution component, and use the Δ -criterion to nucleate new, embryonic dislocation loops. Likewise, experimental studies of the initiation and early stages of nanoscale deformation via slip, twinning or cracking could be designed to exploit the above-mentioned effects of crystallographic orientation, elastic anisotropy and boundary conditions. □

Received 25 February; accepted 22 May 2002; doi:10.1038/nature00865.

- Gerberich, W. W., Venkataraman, S. K., Huang, H., Harvey, S. E. & Kohlstedt, D. L. The injection of plasticity by millineutron contacts. *Acta Metal. Mater.* **43**, 1569–1576 (1995).
- Suresh, S., Nieh, T.-G. & Choi, B. W. Nanoindentation of copper thin films on silicon substrates. *Scripta Mater.* **41**, 951–957 (1999).
- Kiely, J. D., Jarausch, K. F., Houston, J. E. & Russell, P. E. Initial stages of yield in nanoindentation. *J. Mater. Res.* **15**, 4513–4519 (1999).
- Gouldstone, A., Koh, H.-J., Zeng, K. Y., Giannakopoulos, A. E. & Suresh, S. Discrete and continuous deformation during nanoindentation of thin films. *Acta Mater.* **48**, 2277–2295 (2000).
- Kramer, D. E., Yoder, K. B. & Gerberich, W. W. Surface constrained plasticity: Oxide rupture and the yield point process. *Phil. Mag. A* **81**, 2033–2058 (2001).
- Gouldstone, A., Van Vliet, K. J. & Suresh, S. Nanoindentation: Simulation of defect nucleation in a crystal. *Nature* **411**, 656 (2001).
- Kelchner, C. L., Plimpton, S. J. & Hamilton, J. C. Dislocation nucleation and defect structure during surface indentation. *Phys. Rev. B* **58**, 11085–11088 (1998).
- Zimmerman, J. A., Kelchner, C. L., Klein, P. A., Hamilton, J. C. & Foiles, S. M. Surface step effects on nanoindentation. *Phys. Rev. Lett.* **87**, 165507 (2001).
- Tadmor, E. B., Miller, R. & Phillips, R. Nanoindentation and incipient plasticity. *J. Mater. Res.* **14**, 2233–2250 (1999).
- Shenoy, V. B., Phillips, R. & Tadmor, E. B. Nucleation of dislocations beneath a plane strain indenter. *J. Mech. Phys. Solids* **48**, 649–673 (2000).
- Wang, J., Li, J., Yip, S., Phillpot, S. & Wolf, D. Mechanical instabilities of homogeneous crystals. *Phys. Rev. B* **52**, 12627–12635 (1995).
- Zhou, Z. & Joos, B. Stability criteria for homogeneously stressed materials and the calculation of elastic constants. *Phys. Rev. B* **54**, 3841–3850 (1996).
- Morris, J. W. & Krenn, C. R. The internal stability of an elastic solid. *Phil. Mag. A* **80**, 2827–2840 (2000).
- Hill, R. Acceleration waves in solids. *J. Mech. Phys. Solids* **10**, 1–16 (1962).
- Rice, J. R. in *Theoretical and Applied Mechanics* (ed. Koiter, W. T.) Vol. 1, 207–220 (North-Holland, Amsterdam, 1976).
- Egami, T., Maeda, K. & Vitek, V. Structural defects in amorphous solids. A computer simulation study. *Phil. Mag. A* **41**, 883–901 (1980).
- Ray, J. R. Elastic constants and statistical ensembles in molecular dynamics. *Comput. Phys. Rep.* **8**, 109–152 (1988).
- Erickson, J. L. in *Phase Transformations and Material Instabilities in Solids* (ed. Gurtin, M. E.) 61–78 (Academic, Orlando, 1984).
- Tadmor, E. B., Ortiz, M. & Phillips, R. Quasicontinuum analysis of defects in solids. *Phil. Mag. A* **73**, 1529–1563 (1996).
- ABAQUS Theory Manual Version 6.1 (Hibbitt, Karlsson and Sorensen, Pawtucket, Rhode Island, 2000).
- Ackland, G. J., Bacon, D. J., Calder, A. F. & Harry, T. Computer simulation of point defect properties in dilute Fe-Cu alloy using a many-body interatomic potential. *Phil. Mag. A* **75**, 713–732 (1997).
- Brochard, S., Beauchamp, P. & Grilhe, J. Simulations of dislocation nucleation from atomic size surface steps and grooves. *Mater. Sci. Eng. A* **309–310**, 456–462 (2001).
- Hwu, C. & Fan, C. W. Solving the punch problems by analogy with the interface crack problems. *Int. J. Solids Struct.* **35**, 3945–3960 (1998).
- Erolessi, F. & Adams, J. B. Interatomic potentials from first-principles calculations: the force-matching method. *Europhys. Lett.* **26**, 583–588 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

This work was supported by the Defense University Research Initiative on NanoTechnology (DURINT) on 'Damage- and Failure-Resistant Nanostructured and Interfacial Materials' which is supported at the Massachusetts Institute of Technology by the Office of Naval Research. We thank A. S. Argon for comments. K.J.V.V. acknowledges the National Defense Science and Engineering Graduate Fellowship programme. J.L., T.Z. and S.Y. acknowledge support by Honda R&D, AFOSR, NSF/KDI/DMR, and Lawrence Livermore National Laboratory.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.Y. (e-mail: syip@mit.edu).

Dynamic fracture by large extraterrestrial impacts as the origin of shatter cones

Amir Sagy*, Ze'ev Reches* & Jay Fineberg†

* Institute of Earth Sciences, † The Racah Institute of Physics, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

A large impact by a comet or meteorite releases an enormous amount of energy, which evaporates, melts and fractures the surrounding rocks^{1–4}. Distinctive features of such impacts are 'shatter cones', deformed rocks characterized by hierarchical striated features^{5,6}. Although such features have been used for decades as unequivocal fingerprints of large-body impacts, the process by which shatter cones form has remained enigmatic. Here we show that the distinctive shatter-cone striations naturally result from nonlinear waves (front waves) that propagate along a fracture front^{7–10}. This explains the observed systematic increase of striation angles with the distance from the impact. Shatter-cone networks, typically spanning many scales, can be understood as hierarchical bifurcations of the fracture front, which is generated by the immense energy flux carried by the initial, impact-generated, shock waves. Our quantitative predictions based on this theory are supported by field measurements at the Kentland and Vredefort impact sites. These measurements indicate that shatter cones near to the impact site were formed by fractures propagating at nearly the Rayleigh wave speed of the host rocks, whereas the furthest shatter cones observed (about 40 km from the impact site) were formed by fronts moving more slowly. These results provide insight into impact dynamics as well as dissipative mechanisms in solids subjected to sudden, extremely intense fluxes of energy.

The formidable shock waves generated during large extraterrestrial impacts intensely deform crustal rock^{1–4}. Beyond the near-impact region, where rock evaporation and melting prevail, shock-induced structures are dominant^{3,11}. Some of the most distinct of these structures are shatter cones (Fig. 1a and b), which are observed in nature only at large impact sites. They range in size from a few centimetres to a few metres^{5,11}, and generally occur within hierarchical networks—called 'horse-tail' structures⁵—in which cascades of contiguous cones of monotonously decreasing size are observed (Fig. 1b). Although shatter cones are frequently semiconical, complete cones are rare^{12,13}, and, in many cases, their characteristic striations are observed on nearly planar surfaces¹⁴ (Fig. 1b).

The precise mechanism for shatter-cone formation is unknown. It has been shown that conical shapes could result either from the interaction of shock waves with point inhomogeneities in rocks¹⁵, or from interactions between the main compressive shock and rebound waves¹⁶. However, these models do not explain the dominant features of shatter cones: characteristic striations (Fig. 1a), the 'horse-tail' cone hierarchy (Fig. 1b), and the rarity of complete cones. Here we present a new model for shatter-cone formation that is based on recent developments in the study of dynamic fracturing^{7–10}. This model explains the characteristic features of shatter cones, and its predictions are supported by field observations at the classic impact sites of Vredefort, South Africa, and Kentland, Indiana.

The Kentland impact deformed a thick sequence of Mesozoic carbonate and clastic rocks. A quarry, 1 km² in area and 200 m in depth, in the central, uplifted part of this impact provides large, three-dimensional shatter-cone exposures. Our fieldwork reveals that shatter cones are not separable, isolated objects within the rock mass, but are secondary structures that are generated along the

surfaces of large, coherent fractures. At Kentland, for example, the axes of dozens of shatter cones associated with a single large fracture display a well-organized pattern in which the cone axes are directionally aligned over distances of tens of metres. Individual cones range in shape from triangular markings on planar fracture faces to partly conical surfaces that branch away from the host fracture. Examination of the internal cone structure reveals multiple curved, striated surfaces that are essentially secondary branches of the parent fracture (Fig. 2). We have observed up to five levels of hierarchical branches, ranging in size over three to four orders of magnitude. The 'horse-tail' structure of shatter cones may now be understood as a hierarchical network of secondary 'branched' fractures.

These large fractures that generate shatter cones are consistent with tensile fractures formed as a result of the radially propagating shock waves of the impact. Theoretical¹⁷, computational^{18–20} and experimental work^{21–24} has demonstrated that such impact-generated waves produce large tensile stresses normal to the compressive (radial) direction. These tensile stresses, which are predominantly formed at the waves' trailing edge^{18–20,24}, will generate tensile fractures that propagate outwards from the axis of impact^{19,21–24}. We identify shatter cones with secondary tensile fractures that branch away from the large, primary tensile fractures produced by the impact event. The large-scale coherence induced by the main

fracture explains why shatter-cone axes generally point towards the impact centre^{5,6}. This picture also explains the hierarchical branched structure of shatter cones (Fig. 2), which is typically observed in high-velocity tensile fracture in both laboratory experiments²⁵ and in the field²⁶. Shatter-cone curvature reflects the initially curved profile observed in these (secondary) branched fractures²⁵. The role of shear in shatter-cone formation is secondary. Although a small amount of shear displacement is evident in some shatter cones, no such displacement is observed in most of our samples.

We now turn to the striated surface features that are the hallmark of shatter cones (Figs 1 and 2). Striations are restricted solely to shatter-cone surfaces. Within a given shatter-cone sample, additional (sub-surface) striations may be observed. However, these only occur on well-defined branched fractures that have bifurcated beneath the main fracture (see, for example, Fig. 2a). We now show that shatter-cone striations can also be explained in the framework of dynamic tensile fracture.

Recent work has revealed a type of elastic wave that exists along the one-dimensional front defined by the leading edge of propagating tensile fractures. These waves, termed front waves (FWs), were theoretically predicted to result from the interaction between

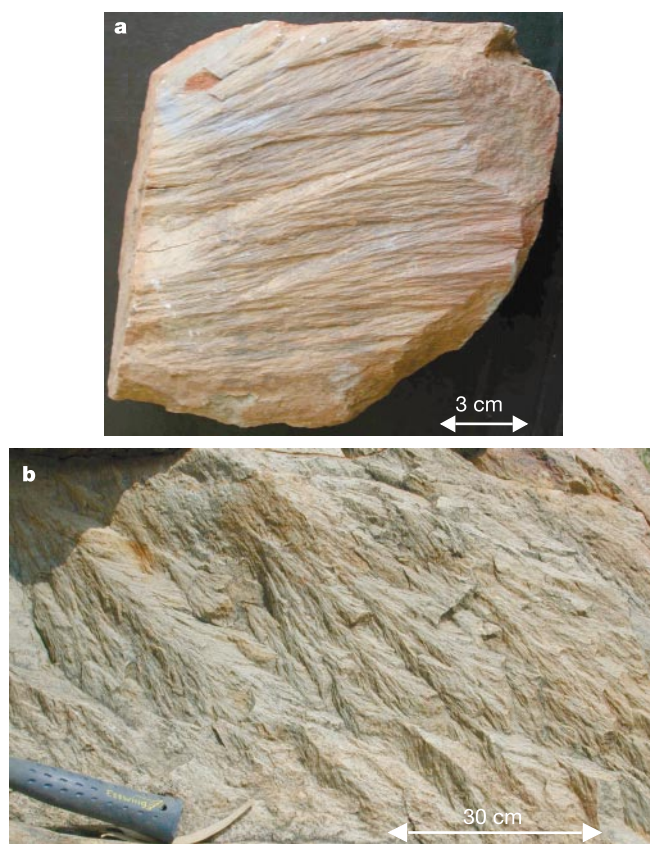


Figure 1 The structure of shatter cones. **a**, A typical quasi-conical surface of a shatter cone on a slate sample. Large numbers of striations (elongated traces with a preferred direction) are observed along shatter-cone surfaces. Pairs of striations originate at points along the shatter-cone surface, and emanate outwards at small angles. Striation depths and widths range from $\sim 10 \mu\text{m}$ to a few millimetres. Scanning electron microscope images reveal that a striation's minimal width corresponds to the grain size of the host rock. **b**, Typical hierarchical, 'horse-tail' structures of a cone-on-cone network on a quasi-planar surface in quartzite. Aligned striations appear on the surface of each cone. Both samples are from the Vredefort impact site.

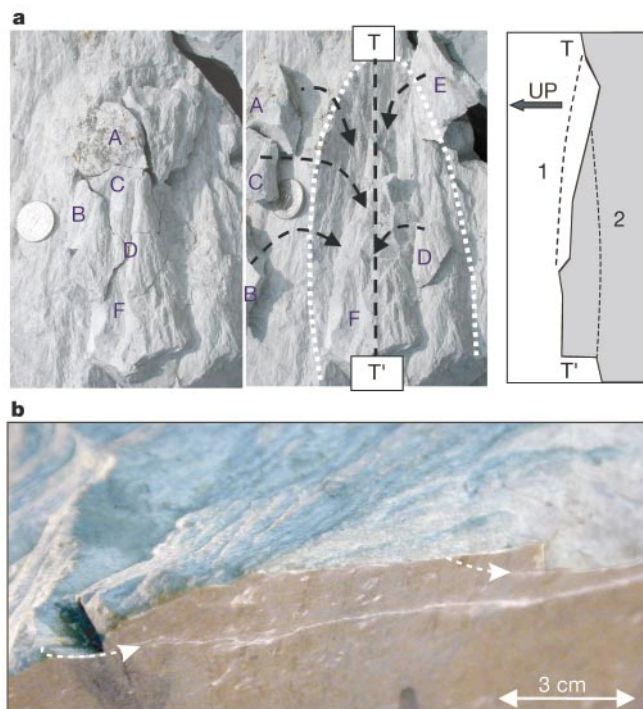


Figure 2 Shatter cones as branched fractures. **a**, Six fragments (labelled A–F) peeled away from the face of a shatter cone reveal a typical branched structure. Left, the fragments in their original positions. Centre, the striated surface of an additional sub-surface cone is clearly visible beneath the removed fragments. This sub-surface cone is formed by a spoon-like branched fracture generated at point T. Right, a crosscut, perpendicular to the surface along the T–T' line, displays the two branched surfaces. The continuous black line indicates the exposed surfaces after removal of the fragments A–F. The dashed line labelled 1 is the initial (upper) surface (before the removal of A–F). The dashed line labelled 2 indicates an additional unexposed branch. The coin in **a** is 22 mm in diameter. **b**, Although shatter-cone surfaces are highly striated, their sub-surface structure generally consists of relatively few branched fractures. An oblique view of a saw-cut shatter cone displays its sub-surface structure. The bright upper area is its external striated surface. Beneath this surface, two branches are observed. White arrows mark the initial branching locations and consequent propagation directions; the irregular shape of the larger branch indicates additional small-scale branching. Note that the striations are found solely on the fracture surfaces (fractographic features) defined by the branches. Both **a** and **b** are dolomite samples from the Kentland site.

material inhomogeneities and fracture fronts^{7–9}. Experiments¹⁰ have shown that FWs, initiated by localized inhomogeneities, create characteristic tracks on the fracture surface (Fig. 3). Pairs of tracks emanate from a local inhomogeneity with a relative angle, α , described by the relation $\cos(\alpha/2) = V/V_{FW}$, where V and V_{FW} are, respectively, the propagation velocities of the fracture front and the FW (Fig. 3). V_{FW} is approximately $(0.96–1.0)V_R$ (refs 7–9), where V_R , the Rayleigh wave speed of the material, is the maximum speed at which a tensile crack can propagate²⁷. This maximal fracture velocity is independent of both the intensity of the tensile stress driving the fracture and the velocity of the shock wave responsible for the loading. Angles greater than $\alpha = 105^\circ$ ($V < 0.6V_R$; ref. 25) have been observed in laboratory experiments (Fig. 3). Because the stress needed to drive a crack increases precipitously beyond this velocity, smaller angles, corresponding to higher fracture velocities, are nearly impossible to achieve under laboratory conditions.

We now postulate that each striation on a shatter-cone surface is a FW track formed along the rapidly propagating fracture front generated by the impact. This would predict that: (1) the angles between striations at a given shatter-cone site should fall within a narrow range, because α is solely determined by V/V_{FW} ; and (2) α should systematically increase with the distance from the impact centre, as the fracture velocity will decrease with the decreasing stress away from the centre.

The Vredefort impact in South Africa (Fig. 4a) provides an excellent test of this hypothesis. The Vredefort site, created about two billion years ago, is one of the largest impact structures on Earth, about 200 km in diameter, and countless shatter cones are exposed within a belt around its centre^{4,14,28}. We analysed shatter cones at 22 sites distributed in the northern section of the Vredefort structure, at distances ranging from 14 km to 37.5 km from its centre (Fig. 4a). The sites are located on exposures of granites, quartzite, slates, chert and andesites, and cover most of the range of shatter-cone distribution in the Vredefort site. Semiconic and semiplanar striated surfaces are well-developed in the fine-grain rocks (slates and quartzite), and are less developed in the coarse-

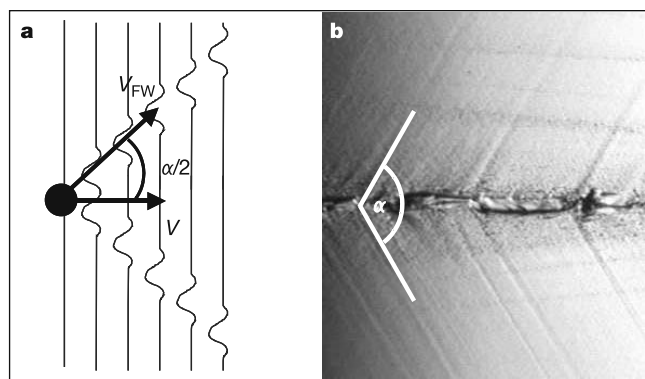


Figure 3 Fracture front waves, FWs, are generated by the interaction of a dynamic fracture front with a localized inhomogeneity. **a**, A schematic presentation of FWs generated by an inhomogeneity (filled circle). Shown are sequential locations of the leading edge of a fracture (a 'fracture front') propagating from left to right with velocity V . Each line marks the position of the fracture front at successive, evenly spaced, time intervals. Two counter-propagating FWs, emitted at the inhomogeneity, propagate along the front with velocity V_{FW} with respect to their initiation point. FWs create structure within the fracture plane that can be observed as characteristic tracks on the fracture surface. The propagation velocity, V , can be determined from the angle α by¹⁰: $\cos(\alpha/2) = V/V_{FW}$. **b**, Photograph of a fracture surface in glass, with FW tracks generated by localized inhomogeneities induced by the spontaneous formation of small branched fractures²⁵ ('micro-branches'). As indicated by $\alpha \approx 105^\circ$, this fracture propagated from left to right at $V \approx 0.65V_{FW}$.

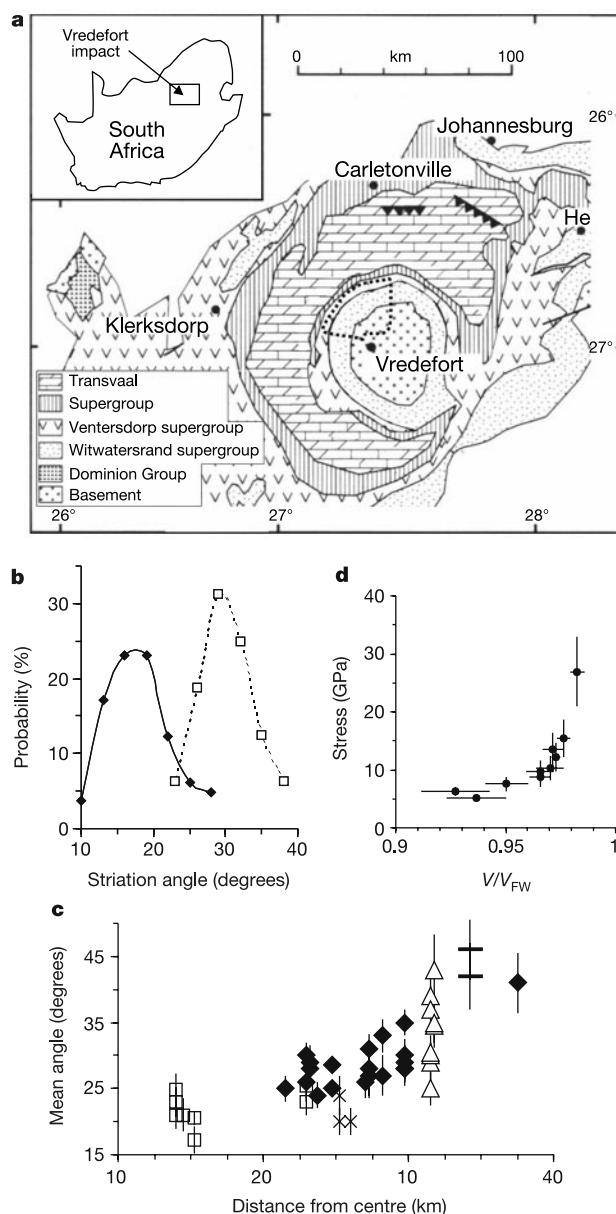


Figure 4 Shatter-cone measurements at the Vredefort impact site. **a**, A simplified map of the Vredefort impact structure³⁰. The locations of the 22 sites at which shatter cones were studied are denoted by the dotted lines. Their present location is a cumulative result of three deformation stages: intra-impact excavation, post-impact relaxation and regional erosion over the last 2 billion years⁴. **b**, In a given shatter cone, the angles between striations are narrowly distributed. Histograms of striation angles, α , measured in slate (diamond symbols) and quartzite (squares) samples, found at different sites are presented. **c**, The mean striation angles, α , systematically increase with the distance from the impact site. Shown are mean striation angles corresponding to granite (squares), quartzite (filled diamonds), slates (crosses), andesite (triangles), and chert (bars). Error bars are 1 s.d. of the measured angles at each site. We ascribe the scatter of α at a given site to variance of the mechanical properties of the rocks. **d**, The relation between the fracture propagation velocity, calculated by $\cos(\alpha/2) = V/V_{FW}$, and the local value of the calculated compressive stresses⁴ induced by the Vredefort impact event. Note the sharp, highly nonlinear, increase of the stress needed to drive propagation velocities that approach the maximal limit of $V_R \approx V_{FW}$. Tensile stresses in the wake of the compressive shock actually drive the fractures. These, which have not been explicitly calculated for the Vredefort event, are of the same order of magnitude as compressive components^{17–19}. Our evaluation of the impact stresses at the measurement sites is based on the numerical modelling of the first two deformation stages for the Vredefort crater by Turtle and Pierazzo⁴. Their best model was calculated for a 10-km-diameter impactor having an impact velocity of 20 km s^{-1} . We assume 5–10 km of erosion^{4,30}.

grained rocks (granite). We used samples from the 22 sites to measure the angle α between pairs of striations. At every site, well-defined striation angles were measured on a number of samples. As predicted, the angles measured within an individual sample are relatively constant with a typical standard deviation of 5° (Fig. 4b). In addition, the mean striation angles systematically increase from 17° to 46° with the distance from the impact centre (Fig. 4c). Thus, both the narrow angular distributions and the systematic increase of mean angles are consistent with our postulate that the striations are formed by front waves.

Using $\cos(\alpha/2) = V/V_{FW} \approx V/V_R$, we calculated the fracture velocities from the measured angles (Fig. 4c). The specific values of V_R in the samples are immaterial, because the angle α measures only the normalized fracture velocity, V/V_{FW} . We find that V approaches $0.98V_R$ at approximately 15 km from the impact centre, and decreases to $\sim 0.9V_R$ at 35–40 km from the centre. Thus, shatter cones are observed within the narrow range of fracture velocities above $0.9V_R$. Using the calculated spatial distribution of the stress applied by the shock wave at Vredefort⁴, Fig. 4d indicates a precipitous rise of the applied stresses with fracture velocity. This increase is consistent with laboratory experiments²⁹ at lower velocities, and amply demonstrates why mean fracture velocities approaching V_R are never observed in the laboratory. At the larger striation angles, we find that shatter cones are not as easily recognized, as clear striations are less frequently observed, possibly because their amplitudes are reduced at lower energy flux levels.

We have shown that shatter cones are branched tensile fractures. Shatter-cone striations are the preserved tracks of fracture front waves. Analysis of the striations shows that shatter cones develop only at extreme propagation velocities, between $0.9V_R$ and the maximal permitted velocity of V_R . The angles of the striations (α), which are shown to increase systematically with the distance from the impact, reflect both the stresses and the energy flux driving the fracture at a given site, and may be used as a general tool to evaluate extreme local stresses in the field. Our results also demonstrate that such rapid fracture propagation in intact solids necessitates extremely high-energy fluxes, supplied naturally only by large impactors. □

Received 5 February; accepted 10 June 2002; doi:10.1038/nature00903.

- Melosh, H. J. Impact ejection, spallation, and the origin of meteorites. *Icarus* **59**, 234–260 (1984).
- Melosh, H. J. *Impact Cratering: A Geologic Process* (Oxford Univ. Press, New York, 1989).
- French, B. M. *Traces of Catastrophe. Handbook of Shock-metamorphic Effects in Terrestrial Meteorite Impact Structures* (Lunar and Planetary Institute, Houston, 1998).
- Turtle, E. P. & Pierazzo, E. Constraints on the size of the Vredefort impact crater from numerical modeling. *Meteorit. Planet. Sci.* **33**, 483–490 (1998).
- Dietz, R. S. in *Shock Metamorphism of Natural Materials* (eds French, B. M. & Short, N. M.) 267–285 (Mono Book Corp., Baltimore, 1969).
- Sharpton, V. L., Dressler, B. O., Herrick, R. R., Schnieders, B. & Scott, J. New constraints on the Slate Islands impact structure, Ontario, Canada. *Geology* **24**, 851–854 (1996).
- Morrissey, J. W. & Rice, J. R. Crack front waves. *J. Mech. Phys. Solids* **46**, 467–487 (1998).
- Morrissey, J. W. & Rice, J. R. Perturbative simulations of crack front waves. *J. Mech. Phys. Solids* **48**, 1229–1251 (2000).
- Ramanathan, S. & Fisher, D. S. Dynamics and instabilities of planar tensile cracks in heterogeneous media. *Phys. Rev. Lett.* **79**, 877–880 (1997).
- Sharon, E., Cohen, G. & Fineberg, J. Propagating solitary waves along a rapidly moving crack front. *Nature* **410**, 68–71 (2001).
- Milton, D. J. in *Impact and Explosion Cratering: Planetary and Terrestrial Implications* (eds Roddy, D. J., Pepin, R. O. & Merrill, R. B.) 703–714 (Pergamon, New York, 1977).
- Roy, D. W. Shatter cone geometry and description procedure. *Tectonophysics* **60**, T37–T42 (1979).
- Albat, H. M. & Mayer, J. J. Megascopic planar shock fractures in the Vredefort Structure; a potential time marker? *Tectonophysics* **162**, 265–276 (1989).
- Nicolaysen, L. O. & Reimold, W. U. Vredefort shatter cones revisited. *J. Geophys. Res.* **B104**, 4911–4930 (1999).
- Johnson, G. P. & Talbot, R. J. *A Theoretical Study of the Shock Wave Origin of Shatter Cones* Thesis (Air Force Inst. Technol., Wright-Patterson Air Force Base).
- Gash, P. J. S. Dynamic mechanism for the formation of shatter cones. *Nature* **230**, 32–35 (1971).
- Shibuya, T. & Nakahara, I. The semi-infinite body subjected to a concentrated impact load on the surface. *Bull. Jpn Soc. Mech. Eng.* **11**, 983–992 (1968).
- Asphaug, E. et al. Mechanical and geological effects of impact cratering on Ida. *Icarus* **120**, 158–184 (1996).
- Camacho, G. T. & Ortiz, M. Computational modelling of impact damage in brittle materials. *Int. J. Solids Struct.* **33**, 2899–2938 (1996).

- Melosh, H. J., Ryan, E. V. & Asphaug, E. Dynamic fragmentation in impacts — Hydrocode simulation of laboratory impacts. *J. Geophys. Res.* **E97**, 14735–14759 (1992).
- Field, J. E. Brittle fracture: its study and application. *Contemp. Phys.* **12**, 1–31 (1971).
- Ahrens, T. J. & Rubin, A. M. Impact-induced tensional failure in rock. *J. Geophys. Res.* **E98**, 1185–1203 (1993).
- Polansky, C. A. & Ahrens, T. J. Impact spallation experiments; fracture patterns and spall velocities. *Icarus* **87**, 140–155 (1990).
- Arakawa, M., Shirai, K. & Kato, M. Shock wave and fracture propagation in water ice by high velocity impact. *Geophys. Res. Lett.* **27**, 305–308 (2000).
- Sharon, E. & Fineberg, J. Microbranching instability and the dynamic fracture of brittle materials. *Phys. Rev. B* **54**, 7128–7139 (1996).
- Sagy, A., Reches, Z. & Roman, I. Dynamic fracturing: field and experimental observations. *J. Struct. Geol.* **23**, 1223–1239 (2001).
- Freund, L. B. *Dynamic Fracture Mechanics* (Cambridge Univ. Press, Cambridge, 1990).
- Henkel, H. & Reimold, W. U. Integrated geophysical modelling of a giant, complex impact structure; anatomy of the Vredefort Structure, South Africa. *Tectonophysics* **287**, 1–20 (1998).
- Dally, J. W. Dynamic photoelastic studies of fracture. *Exp. Mech.* **19**, 349–361 (1979).
- Therriault, A. M., Grieve, R. A. F. & Reimold, W. U. Original size of the Vredefort Structure; implications for the geological evolution of the Witwatersrand Basin. *Meteorit. Planet. Sci.* **32**, 71–77 (1997).

Acknowledgements

We thank W. U. Reimold and E. G. Charlesworth for information and hospitality at the Vredefort site; the Rogers Group's Newton County quarry for their hospitality at the Kentland site; and G. Cohen for assistance. This work was supported by the United States-Israel Binational Fund.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.F. (e-mail: jay@vms.huji.ac.il).

Climate-mediated energetic constraints on the distribution of hibernating mammals

Murray M. Humphries*†, Donald W. Thomas‡ & John R. Speakman*§

* Department of Zoology, University of Aberdeen, Aberdeen AB24 3TZ, UK

† Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

‡ Département de Biologie, Université de Sherbrooke, Sherbrooke, Québec J1K 2R1, Canada

§ Rowett Research Institute, Bucksburn, Aberdeen AB21 9BS, UK

To predict the consequences of human-induced global climate change, we need to understand how climate is linked to biogeography¹. Energetic constraints are commonly invoked to explain animal distributions, and physiological parameters are known to vary along distributional gradients². But the causal nature of the links between climate and animal biogeography remain largely obscure^{2,3}. Here we develop a bioenergetic model that predicts the feasibility of mammalian hibernation under different climatic conditions. As an example, we use the well-quantified hibernation energetics of the little brown bat (*Myotis lucifugus*) to parameterize the model⁴. Our model predicts pronounced effects of ambient temperature on total winter energy requirements, and a relatively narrow combination of hibernaculum temperatures and winter lengths permitting successful hibernation. Microhabitat and northern distribution limits of *M. lucifugus* are consistent with model predictions, suggesting that the thermal dependence of hibernation energetics constrains the biogeography of this species. Integrating projections of climate change into our model predicts a pronounced northward range expansion of hibernating bats within the next 80 years. Bioener-

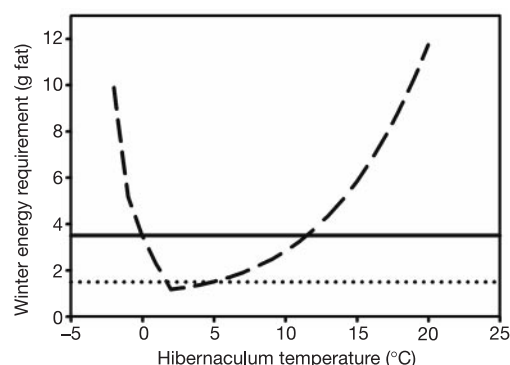


Figure 1 The effect of hibernaculum temperature on total winter energy requirements of *M. lucifugus* (dashed line), based on a winter length of 193 days¹⁰. The horizontal lines indicate approximate maximum (solid line) and minimum (dotted line) hibernation fat stores⁹.

genetics can provide the simple link between climate and biogeography needed to predict the consequences of climate change.

Explanation of the geographic distribution of organisms is a central aim of ecology⁵. Climate is clearly involved, because distribution limits frequently match climatic features such as seasonal minimum or maximum temperature isotherms⁶. Nevertheless, the mechanisms that underlie these patterns remain obscure^{2,3,7}. In many cases, the ability of species to occupy a given environment may be determined by seasonal energetic bottlenecks within the annual cycle⁸. In these situations, explanation of distribution limits using climate-based energetic models may be much simpler than previously recognized³. For example, over-winter survival of hibernating mammals should depend primarily on the size of their energy reserves at the onset of hibernation, the rate at which the energy store is depleted during winter, and the length of the winter. If the size of the reserve is less than the rate of depletion times the length of winter, the hibernator will not survive. Here, we model the quantitative relationship between ambient temperature and energy

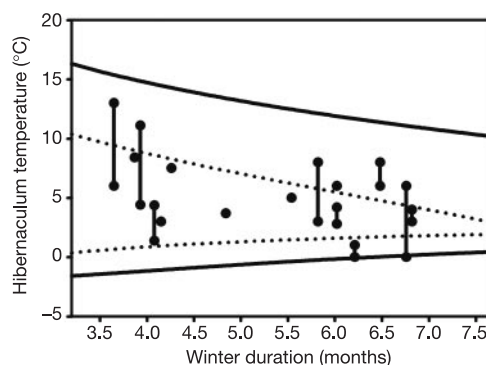


Figure 2 Comparison of observed hibernaculum temperatures of *M. lucifugus* with those predicted to be suitable by the model. Combinations of winter length and hibernaculum temperatures above the upper solid line or below the lower solid line are predicted to require more than the maximum fat stores (3.5 g)⁹ to survive winter. Combinations above or below the dashed lines require more than minimum fat stores (1.5 g)⁹. Circles represent recorded hibernaculum temperatures of *M. lucifugus* at different geographical locations, and circles connected by a line represent a range of occupied temperatures reported from a single location^{10,19–28}. Winter duration for each location is estimated from the period when nightly minima are below 0 °C. Most hibernacula are characterized by temperatures predicted to permit successful hibernation with minimum fat stores (12 of 15), and all are characterized by temperatures permitting successful hibernation with maximum fat stores (15 of 15). Maximum occupied temperatures are significantly lower than maximum suitable temperatures (Student's paired $t_{15} = 10.8$, $P < 0.0001$) and minimum occupied temperatures are significantly higher than minimum suitable temperatures (paired $t_{15} = 6.1$, $P < 0.0001$).

expenditure during hibernation for a common North American bat, then combine this with estimates of winter length and the maximum size of fat stores to predict under what climatic conditions successful hibernation should be possible.

We focus this modelling approach on *M. lucifugus* because of the availability of empirical data on its hibernation energetics and distribution patterns, and the relative ease of predicting the thermal

Box 1

Temperature and hibernation energetics

The energy expenditure of an endotherm with a normal, elevated body temperature (euthermic), E_{eu} , varies with ambient temperature, T_a , according to a well-described metabolic response curve:

$$E_{eu} = RMR + (T_c - T_a)C_{eu}$$

where RMR is resting metabolic rate, T_c is the lower critical temperature, and C_{eu} is euthermic thermal conductance³⁰. During torpor, metabolic rate, TMR , and body temperature decline with T_a until a lower ambient set-point temperature, $T_{tor-min}$, is reached, after which torpor body temperature is defended (that is, remains constant) and consequently TMR increases. Thus, TMR varies with temperature according to

$$E_{tor} = TMR_{min}Q_{10}^{(T_a - T_{tor-min})/10}, \text{ if } T_a > T_{tor-min}$$

$$E_{tor} = TMR_{min} + (T_{tor-min} - T_a)C_t, \text{ if } T_a \leq T_{tor-min}$$

where Q_{10} represents the change in torpor metabolism resulting from a 10 °C change in T_a , and C_t represents torpor conductance below $T_{tor-min}$.

All mammalian hibernators arouse from torpor at regular intervals during winter, remaining euthermic for a short time before re-entering torpor³⁰. The energetic cost of these arousals E_{ar} is a simple function of the required increase in body temperature from T_{tor} to euthermic

levels, T_{eu} , and the specific heat capacity S of the hibernator's tissues³⁰:

$$E_{ar} = (T_{eu} - T_{tor})S$$

The energetic cost of a complete torpor-arousal cycle, E_{bout} , is then

$$E_{bout} = E_{eu}(t_{eu}) + E_{tor}(t_{tor}) + E_{ar}$$

where t_{eu} and t_{tor} are time spent euthermic and torpid. The length of torpor bouts also varies dramatically with ambient temperature, presumably because of the thermal dependency of TMR ³⁰. Thus, t_{tor} reduces from maximum levels, $t_{tor-max}$, above and below $T_{tor-min}$ in accordance with increases in E_{tor} such that

$$t_{tor} = t_{tor-max}Q_{10}^{(T_a - T_{tor-min})/10}, \text{ if } T_a > T_{tor-min}$$

$$t_{tor} = t_{tor-max} + (T_a - T_{tor-min})k, \text{ if } T_a \leq T_{tor-min}$$

where k is an analytical constant that yields equal values of t_{tor} for a given TMR above and below $T_{tor-min}$.

Finally, total hibernation energy requirements E_{winter} can be predicted for a given winter length t_{winter} according to

$$E_{winter} = \left(\frac{t_{winter}}{t_{bout}} \right) E_{bout}$$

where t_{bout} is the temperature-dependent duration of a torpor-arousal cycle.

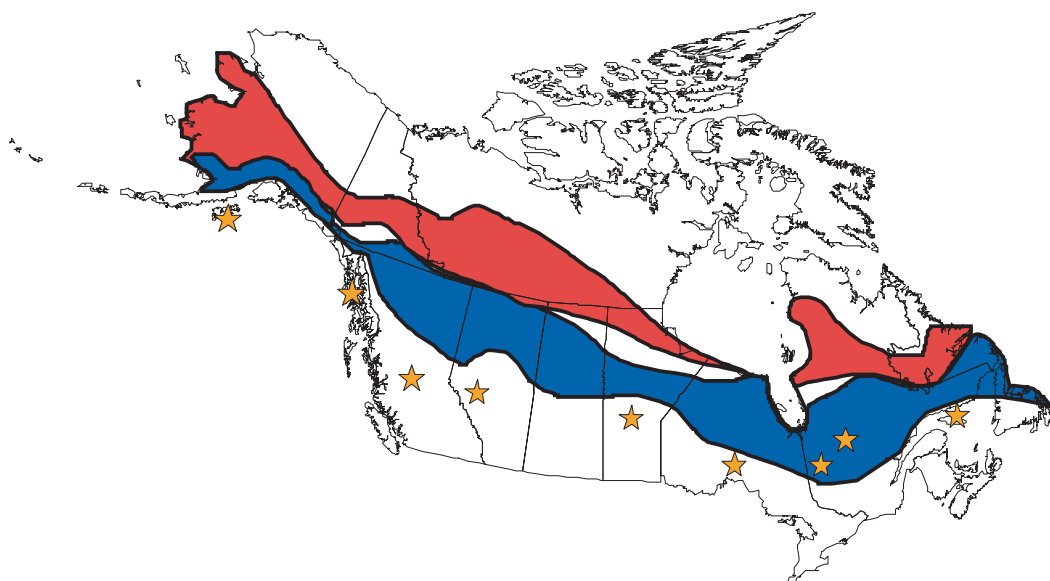


Figure 3 Observed and predicted *M. lucifugus* range distributions in northern North America. Confirmed northern locations of hibernacula (refs 10, 14, 19 and 29, and M. Gauthier, personal communication) are indicated with yellow stars, and the predicted distributional limit of hibernacula is indicated by the blue area; the width of the area represents variation in predictions resulting from cave temperatures exceeding average annual temperature by 0–2 °C, fat stores varying from 1.5 to 3.5 g, and all other parameters varying by up to $\pm 5\%$. Thus, the lower margin of the blue zone indicates the limit beyond which to the south almost all bats could successfully hibernate, and the upper margin indicates the limit beyond which to the north almost no bats could successfully

hibernate. Observed hibernacula approach and only marginally exceed the lower margin of this zone and there is a significant positive correlation between the latitude of observed hibernacula and the lower margin of this zone (predictions generated for each hibernaculum longitude, $n = 9$, $r = 0.93$, $P < 0.001$; intercept not significantly different from 0). The red area represents the predicted 2080 winter range limit of *M. lucifugus* due to global warming. Climate projections are based on the Hadley Centre coupled global climate model (HadCM2—GAX)¹². The width of the red zone represents the same as does the blue zone.

conditions under which it hibernates. This species has a geographic range that extends into northern temperate and subarctic regions where individuals do not feed during winter⁴ and instead rely completely on stored body fat during a prolonged hibernation period⁹. Because of morphological constraints and costs associated with fat storage, pre-hibernation fat reserves rarely exceed 40% of total body mass in hibernating mammals. For *M. lucifugus* in particular, the combination of aerial feeding (limiting the maximum size of a fat store), small body size (resulting in high mass-specific metabolic rate) and long winters in the northern part of its range impose severe constraints on the amount of energy available for hibernation⁹. Using a modelling approach applicable to all hibernating mammals, we estimate hibernation energy and fat requirements of *M. lucifugus* on the basis of the temperature dependency of euthermic metabolism, torpor metabolism, and the frequency and metabolic costs of arousals (Box 1). Because this species is an obligate insectivore that hibernates in caves and abandoned mines, both hibernaculum temperature and winter length can be readily estimated from climatic data (see Methods).

Our model predicts that hibernation energy requirements will be minimized at 2 °C and increase sharply with either an increase or decrease in hibernaculum temperature (Fig. 1). Energy requirements increase dramatically below 2 °C because in this temperature range, the energy costs of both torpor and euthermia increase as temperature decreases. Requirements increase more moderately above 2 °C because above this temperature the costs of torpor increase, but the costs of arousal and euthermia both decrease. Given a winter length of 193 days in the northern portion of the range of *M. lucifugus*¹⁰, the temperature dependency of hibernation energetics should severely limit the range of temperatures where successful hibernation is possible. Hibernating at 2 °C for 193 days requires only 1.2 g of fat, but temperatures only 3 °C lower or 10 °C

higher increase energy requirements threefold, exceeding the observed 3.5 g upper limit to fat stores⁹ (Fig. 1). The validity of our model can be evaluated by comparing predicted combinations of winter length and hibernaculum temperature for which successful hibernation should be possible, with reported hibernacula temperatures of *M. lucifugus* at different latitudes. These empirical studies show that bats hibernate across a relatively wide temperature range within a geographic location¹¹, but that this range does not extend to regions predicted unsuitable by the model (Fig. 2).

The pronounced increase in hibernation energy requirements at low ambient temperatures, combined with constraints on the size of fat reserves at the onset of hibernation and the length of the hibernation season, permits application of our model to predict the northern biogeographical limit for *M. lucifugus* hibernation. So applied, our model predicts that the minimum fat stores typically accumulated by this species⁹ are inadequate for successful hibernation throughout the northern portions of the Canadian provinces, and that maximum accumulated stores are inadequate throughout most of Alaska and the Canadian territories (Fig. 3). Confirmed northern hibernacula of *M. lucifugus* extend to mid-latitude regions of Canadian provinces and maritime-influenced areas of Alaska, where our model predicts successful hibernation by most individuals, but not into higher latitudes and continental regions, where our model predicts that successful hibernation is energetically impossible (Fig. 3). Energetic constraints at higher latitudes should be especially severe for juveniles, owing to their limited capacity to grow and fatten during a short active season⁹.

The correspondence between predicted and observed hibernaculum microclimates (Fig. 2) and between predicted and observed northern hibernaculum localities (Fig. 3) provides strong evidence that the distribution of *M. lucifugus* is constrained by thermal effects on hibernation energetics. This mechanistic link between tempera-

ture and biogeography permits prediction of the future impacts of climate change. Current global climate models predict a 6–8 °C increase in average annual temperature throughout most of northern North America within the next 80 years¹². On the basis of our model (see Methods), this climatic warming will result in a pronounced northward expansion of *M. lucifugus* (Fig. 3).

The success of this simple model, rooted in the fundamental energetics of the species, in predicting biogeographical distributions, demonstrates how the study of energetics can link climate to biogeography. Previous studies have demonstrated a correspondence between distributional temperature and several physiological processes², but have been unable to explain how these correspondences generate species range limits⁷. Establishing this link in *M. lucifugus*, using an approach that can be generalized to many other organisms that experience seasonal energetic bottlenecks, facilitates the use of bioenergetics to explain the present-day range boundaries and to predict the sensitivity of these boundaries to future climate change. The pronounced northward expansion of little brown bats predicted to occur within the next century provides a clear but singular example of the major impacts that human-induced climate change will have on the range limits of endotherms. Enhancing our understanding of the links between bioenergetics and biogeography in a wider diversity of species will improve our capacity to anticipate the biological impacts of global climate change. □

Methods

Model parameters and sensitivity

Parameter values for *M. lucifugus* were obtained from the literature (see Supplementary Information; see Box 1 for parameter definitions). Model sensitivity was first investigated by determining how independent variation in each parameter influenced our estimate of total winter energy requirements (E_{winter}) at $T_a = 0^\circ\text{C}$. The proportion of variation in each parameter reflected in variation in E_{winter} ('sensitivity') ranges from 0.01 to 0.77. For example, a 5% change in the value of $T_{\text{tor-min}}$ (sensitivity = 0.77) would alter our E_{winter} estimate by slightly less than 4%, whereas a 5% change in t_{ar} (sensitivity = 0.01) would alter E_{winter} by less than 0.1%. Sensitivity is highest for parameters related to the duration and energetic cost of torpor bouts (see Supplementary Information). The influence of combined variation of all terms was simulated by calculating E_{winter} at $T_a = 0^\circ\text{C}$ using parameter values drawn randomly from a normal distribution with 99% of values within $\pm 5\%$ of the original parameter value. On the basis of 50,000 runs, the average deviation from our original E_{winter} prediction was only $\pm 2\%$ and the maximum was $\pm 10\%$. Finally, the maximum potential deviation in our E_{winter} estimate at 0°C resulting from a combined $\pm 5\%$ error in all parameters was determined analytically and was found to be $\pm 17\%$. This sensitivity analysis ignores potential co-dependence among parameter values and thus will tend to overestimate the model's sensitivity.

Estimating winter length from climatic data

Because *M. lucifugus* is a nocturnal obligate insectivore⁴ and near-freezing air temperatures inhibit insect activity¹³, the minimum length of the hibernation period should approximate the period when average nightly minimum temperatures are less than 0°C (night-time maximum temperatures would typically exceed this minimum by several degrees). Thus, climatological data from weather stations (Canadian Meteorological Centre, Climate and Water Information, Environment Canada (<http://www.msc-smc.gc.ca/climate>); National Climatic Data Centre, National Oceanic and Atmospheric Administration, USA (<http://lwf.ncdc.noaa.gov/oa>)) can be used to predict how the duration of the hibernation period varies for *M. lucifugus* across its wide geographic distribution. Predictions for the onset, termination and total duration of hibernation at different latitudes, derived from our 0°C nightly minimum assumption, correspond well with dates and durations reported in the literature^{9,10,14–16}, with an average absolute error of 5 days (range 0–13 days).

Estimating hibernaculum temperature from climatic data

Although the temperature profiles of caves and abandoned mines vary according to internal geometry, deep cave and mine temperatures generally approximate average annual temperature, whereas regions nearer to the entrance track winter temperatures more closely¹⁷. As a result, although the coldest sites available for hibernation are much less than annual average temperatures, the warmest sites exceed the annual average by only a few degrees, unless geothermally heated caves occur in the region¹⁸. We determined average annual temperature for each region, and assumed that the maximum hibernaculum temperature available in a given region could exceed this average by 0–2 °C; the width of the coloured areas in Fig. 3 incorporates the effect of this temperature range on the predicted northern distribution.

Climate forecasting

We obtained monthly mean and minimum temperature projections for 2020 from a global climate model (Hadley Centre Coupled Model 2 with aerosol simulation mean, HadCM2—GAX)¹² made available by the Canadian Climate Impacts and Scenarios

(<http://www.cics.uvic.ca>) in collaboration with the Canadian Centre for Climate Modelling and Analysis and the Intergovernmental Panel on Climate Change. Because monthly temperature projections were available for a fine-scale georeferenced grid (cell size 2.50° latitude $\times$ 3.75° longitude), cave temperatures and winter lengths could be estimated for each cell, and entered into our model to evaluate the future potential for successful *M. lucifugus* hibernation across northern North America.

Received 21 March; accepted 22 April 2002; doi:10.1038/nature00828.

- Gates, D. M. *Climate Change and Its Biological Consequences* 162–201 (Sinauer, Sunderland, Massachusetts, 1993).
- Johnston, I. A. & Bennett, A. F. (eds) *Animals and Temperature: Phenotypic and Evolutionary Adaptation* (Cambridge Univ. Press, Cambridge, 1996).
- Chown, S. L. & Gaston, K. J. Exploring links between physiology and ecology at macro-scales: the role of respiratory metabolism in insects. *Biol. Rev.* **74**, 87–112 (1999).
- Fenton, M. B. & Barclay, R. M. R. *Myotis lucifugus*. *Mammal. Spec.* **142**, 1–8 (1980).
- Rosenzweig, M. L. *Species Diversity in Space and Time* 8–48 (Cambridge Univ. Press, Cambridge, 1995).
- Root, T. Environmental factors associated with avian distributional boundaries. *J. Biogeogr.* **15**, 489–505 (1988).
- Canterbury, G. Metabolic adaptation and climatic constraints on winter bird distribution. *Ecology* **83**, 946–957 (2002).
- Weiner, J. Physiological limits to sustainable energy budgets in birds and mammals—ecological implications. *Trends Ecol. Evol.* **7**, 384–388 (1992).
- Kunz, T. H., Wrazen, J. A. & Burnett, C. D. Changes in body mass and fat reserves in pre-hibernating little brown bats (*Myotis lucifugus*). *Ecoscience* **5**, 8–17 (1998).
- Fenton, M. B. Population studies of *Myotis lucifugus*. (Chiroptera: Vespertilionidae) in Ontario. *Life Sci. Contr.*, *R. Ont. Mus.* **77**, 1–34 (1970).
- Webb, P. L., Racey, P. & Speakman, J. R. How hot is a hibernaculum? A review of the temperatures at which bats hibernate. *Can. J. Zool.* **74**, 761–765 (1996).
- Johns, T. C. et al. The second Hadley Centre couple ocean-atmosphere GCM: model description, spinup and validation. *Clim. Dyn.* **13**, 103–134 (1997).
- Rainey, R. C. in *Insect Flight* (ed. Rainey, R. C.) 75–112 (Blackwell Scientific, Oxford, 1976).
- Schowalter, D. B. Swarming, reproduction, and early hibernation patterns of *Myotis lucifugus* and *M. volans* in Alberta, Canada. *J. Mammal.* **61**, 347–350 (1980).
- Whitaker, J. O. & Rissler, L. J. Winter activity of bats at a mine entrance in Vermillion County, Indiana. *Am. Midl. Nat.* **127**, 52–59 (1992).
- Raesley, R. L. & Gates, J. E. Winter habitat selection by north temperate cave bats. *Am. Midl. Nat.* **118**, 15–31 (1986).
- Dwyer, P. D. Temperature regulation and cave-dwelling in bats: an evolutionary perspective. *Mammalia* **35**, 424–455 (1971).
- Bell, G. P., Bartholomew, G. A. & Nagy, K. A. The roles of energetics, water economy, foraging behaviour, and geothermal refugia in the distribution of the bat, *Macrotus californicus*. *J. Comp. Physiol. B* **156**, 441–450 (1986).
- Parker, D. L., Lawhead, B. E. & Cook, J. A. Distributional limits of bats in Alaska. *Arctic* **50**, 256–265 (1997).
- Twente, J. W. Environmental problems involving the hibernation of bats in Utah. *Proc. Utah Acad. Sci. Arts Lett.* **37**, 67–71 (1960).
- Menaker, M. Hibernation-hypothermia: an annual cycle of response to low temperature in the bat *Myotis lucifugus*. *J. Cell. Comp. Physiol.* **59**, 163–173 (1962).
- Pearson, E. W. Bats hibernating in silica mines in southern Illinois. *J. Mammal.* **43**, 27–33 (1962).
- Davis, W. H. & Hitchcock, H. B. Notes on sex ratios of hibernating bats. *J. Mammal.* **45**, 475–476 (1964).
- Henshaw, R. E. & Folk, G. E. Jr Relation of thermoregulation to seasonally changing microclimate in two species of bats (*Myotis lucifugus* and *M. sodalis*). *Physiol. Zool.* **39**, 223–236 (1966).
- McManus, J. J. Activity and thermal preference of the little brown bat, *Myotis lucifugus*, during hibernation. *J. Mammal.* **55**, 844–846 (1974).
- Nagorsen, D. W. Records of hibernating big brown bats (*Eptesicus fuscus*) and little brown bats (*Myotis lucifugus*) in Northwestern Ontario. *Can. Field Nat.* **94**, 83–85 (1980).
- Brack, V. Jr & Twente, J. W. The duration of the period of hibernation of three species of vespertilionid bats. I. Field studies. *Can. J. Zool.* **63**, 2952–2954 (1985).
- Thomas, D. W. The physiological ecology of hibernation in vespertilionid bats. *Symp. Zool. Soc. Lond.* **67**, 233–244 (1995).
- Nagorsen, D. W. et al. Winter bat records for British Columbia. *Northwest. Nat.* **74**, 61–66 (1993).
- Speakman, J. R. & Thomas, D. W. in *Bat Biology* (eds Kunz, T. H. & Fenton, M. B.) (Univ. Chicago Press, Chicago, in the press).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank J. Umbanhowar for assistance with model sensitivity analysis as well as D. Kramer and K. McCann for comments on the presentation. This research was supported through a National Sciences and Engineering Research Council of Canada (NSERC) postdoctoral fellowship to M.M.H., and an NSERC research grant to D.W.T.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.M.H. (e-mail: murray_humphries@hotmail.com).

An antioxidant function for DMSP and DMS in marine algae

W. Sunda*, D. J. Kieber†, R. P. Kiene‡ & S. Huntsman*

* Beaufort Laboratory, National Oceanic and Atmospheric Administration, Beaufort, North Carolina 28516, USA

† State University of New York, College of Environmental Science and Forestry, Chemistry Department, 1 Forestry Drive, Syracuse, New York 13210, USA

‡ University of South Alabama, Department of Marine Sciences, Mobile, Alabama 36688, USA

The algal osmolyte dimethylsulphoniopropionate (DMSP) and its enzymatic cleavage product dimethylsulphide (DMS) contribute significantly to the global sulphur cycle^{1–3}, yet their physiological functions are uncertain⁴. Here we report results that, together with those in the literature^{5,6}, show that DMSP and its breakdown products (DMS, acrylate, dimethylsulphoxide, and methane sulphinic acid) readily scavenge hydroxyl radicals and other reactive oxygen species, and thus may serve as an antioxidant system, regulated in part by enzymatic cleavage of DMSP. In support of this hypothesis, we found that oxidative stressors, solar ultraviolet radiation⁷, CO₂ limitation⁸, Fe limitation, high Cu²⁺ (ref. 9) and H₂O₂ substantially increased cellular DMSP and/or its lysis to DMS in marine algal cultures. Our results indicate direct links between such stressors and the dynamics of DMSP and DMS in marine phytoplankton, which probably influence the production of DMS and its release to the atmosphere. As oxidation of DMS to sulphuric acid in the atmosphere provides a major source of sulphate aerosols and cloud condensation nuclei³, oxidative stressors—including solar radiation and Fe limitation—may be involved in complex ocean–atmosphere feedback loops that influence global climate and hydrological cycles^{1,2}.

DMSP occurs at high cellular concentrations (100–400 mmol l^{–1}) in many marine algae, including prymnesiophytes and dinoflagellates, and thus functions as an osmolyte in these algae. DMSP has also been proposed to serve as a cryoprotectant in polar algae and as a grazing deterrent via its cleavage to acrylate, although its overall physiological functions remain unclear⁴. Its cellular concentration increases with light in many algal species^{10,11}, an effect not readily explained by existing proposed functions. In laboratory experiments, we found that DMSP reacts rapidly with the hydroxyl radical ([•]OH), and thus could serve as an effective cellular scavenger of this harmful radical (Table 1). DMSP's enzymatic cleavage products, DMS and acrylate, are even more effective at scavenging [•]OH, as are the DMS oxidation products dimethylsulphoxide (DMSO) and methane sulphinic acid (MSNA) (Table 1). Calculations suggest that, taken together, these molecules constitute an antioxidant system, which could be more effective at scavenging [•]OH in high-DMSP algae than other well-recognized antioxidants, such as ascorbate and glutathione (Table 1). An antioxidant function would explain the observed increase in algal DMSP concentrations at high light^{10,11}, as [•]OH is produced as a by-product of photosynthesis.

All five proposed antioxidant compounds—DMSP, acrylate, DMS, DMSO and MSNA (Table 1)—have unique chemical and physical properties that the cell can exploit. The metabolically compatible zwitterion DMSP is lysed to DMS and acrylate via the enzyme DMSP lyase, whose cellular function and regulation are not well understood¹². Acrylate and DMS are ~20 and ~60 times more reactive towards [•]OH than is DMSP (Table 1), and DMS is highly reactive toward singlet oxygen¹³. Acrylate is charged at physiological pHs, and like DMSP, cannot penetrate into lipid membranes; but uncharged DMS molecules can readily diffuse into photosynthetic

membranes, an important site of harmful lipid peroxidation reactions. Thus, DMSP lysis should substantially increase the antioxidant protection in both aqueous and lipid membrane phases within the cell. This increased protection could be modulated by DMSP lyase, providing an important metabolic function for this enzyme.

Much of the released DMS should oxidize to DMSO or other oxidized sulphur species, while the remainder would diffuse from the cell. The DMSO produced from [•]OH oxidation of DMS or DMSP (see Methods) is much more hydrophilic than DMS, retarding its loss across cell membranes and allowing it to accumulate at high cellular concentrations^{14–16}. These high concentrations, combined with a high reactivity toward [•]OH, makes DMSO an effective antioxidant (Table 1), as has been proposed previously¹⁵ and has been directly observed in isolated chloroplasts⁶. DMSO oxidation by [•]OH produces the water-soluble antioxidant MSNA, which can further react with and scavenge [•]OH¹⁷ (Table 1). Thus, enzymatic lysis of DMSP gives rise to four water- or lipid-soluble antioxidant scavengers of [•]OH radicals and at least one scavenger (DMS) of singlet oxygen. We suggest that this multiplicative effect combined with high cellular DMSP concentrations make DMSP and its lysis and oxidation products a highly effective antioxidant system.

Organisms generally acclimate to oxidative stress by up-regulation of antioxidant systems^{7–9}. Thus, under our hypothesis, phytoplankton should respond to higher levels of oxidative stress by increased DMSP synthesis and lysis to DMS. This indeed is what we observed for the diatom *Thalassiosira pseudonana* under CO₂ limitation, a known oxidative stressor⁸, and Fe limitation, shown here to also oxidatively stress cells. A nutrient-sufficient control culture had a high specific growth rate (1.45 d^{–1}), high cellular chlorophyll *a*, low cellular DMSP (~1 mmol l^{–1}), low DMS to cell-volume ratios, and low cellular activity of the antioxidant enzyme ascorbate peroxidase (APX; Table 2). But under CO₂ limitation and Fe limitation, we observed 20- to 60-fold increases in cellular DMSP, increased DMS/cell-volume ratios, and increased APX activities, in conjunction with large decreases in growth rate and chlorophyll *a*. APX is the main chloroplast enzyme for removal of H₂O₂, a reactive by-product of photosynthesis and a major precursor for [•]OH radical formation^{18,19}. As antioxidant systems are up-regulated under increased oxidative stress, the significant increase in APX activity/chlorophyll *a* ratios under CO₂ limitation (*P* < 0.001, *t*-test) and Fe limitation (*P* < 0.01) indicate that both increased oxidative stress within the chloroplast. The increase in APX activity is noteworthy as APX contains Fe, and would be expected to decrease under Fe limitation. But under Fe limitation, there is a decrease in Fe-containing components of the photosynthetic electron transport chain (for example, the cytochrome *b₆/f* complex) relative to

Table 1 Computed *in vitro* half-lives for [•]OH radical

Compound	Concentration (mmol l ^{–1})	Rate constant* (M ^{–1} s ^{–1})	Ref.	[•] OH half-life (ns)
DMSP	150–450†	3 × 10 ⁸	This work	5–15
Acrylate	~1–40?	5.6 × 10 ⁹	Ref. 5	3–124
DMS	~1–40?	1.9 × 10 ¹⁰	Ref. 5	0.9–36
DMSO	30–90†	6.6 × 10 ⁹	Ref. 5	1.2–3.5
MSNA	?	9 × 10 ⁹	This work	?
Glutathione	2.4‡	1.4 × 10 ¹⁰	Ref. 5	21
Ascorbate	1–6§	1.1 × 10 ¹⁰	Ref. 5	11–63

Data are shown for a range of concentrations of DMSP and its breakdown products, and of antioxidants (glutathione and ascorbate) in high-DMSP algae (for example, *E. huxleyi*).

*Rate constant for reaction of each compound with [•]OH at 25 °C; pH conditions are 7.5 for this work, and 7.0 for ref. 5.

†Approximate DMSP range that we observed in *E. huxleyi*. The cellular DMSO concentration was computed by dividing the cell DMSP concentration by the DMSP/DMSO ratio (~5) found in near-surface oceanic plankton¹⁶.

‡Measured value in *E. huxleyi* (CCMP373) cultures²⁹.

§Estimated average range for phytoplankton based on the reported range per dry weight of algal cells (5–28 mmol g^{–1}; ref. 30), an assumed algal dry to wet weight ratio of 0.8, and a cell density of 1.03 g cm^{–3} (the value for sea water).

Table 2 Effect of CO₂ limitation and Fe limitation in *T. pseudonana* cultures

Treatment	Day	Growth rate (d ⁻¹)	Cell vol. (μl l ⁻¹)*	pH	[CO ₂] (μmol l ⁻¹)	Cell chl <i>a</i> (mmol l ⁻¹)†	Cell DMSP (mmol l ⁻¹)‡	DMS (mmol l ⁻¹)‡	APX act. (per l CV)‡	APX act. (per mol chl)‡
Control	9	1.45	6.6	ND§	ND	4.0	0.88 ± 0.02	0.08 ± 0.1	118 ± 37	29 ± 9
	12	1.45	7.8	8.42	5.1	4.4 ± 0.2	1.46 ± 0.01	0.11 ± 0.0	99 ± 24	23 ± 6
	13	1.45	4.9	8.32	7.0	3.9 ± 0.0	0.85 ± 0.01	0.0	93 ± 18	24 ± 5
CO ₂ -limited	6	0.31	35.6	9.00	0.69	3.0 ± 0.1	18.9 ± 0.2	0.42 ± 0.01	255 ± 13	84 ± 4
	7	0.31	37.8	9.10	0.47	2.9 ± 0.0	13.5 ± 0.2	0.32	209 ± 20	72 ± 7
	10	0.15	39.8	9.26	0.25	1.9 ± 0.0	59.6 ± 1.7	0.59 ± 0.01	305 ± 23	160 ± 12
Fe-limited	20	0.18	46.5	9.40	0.14	1.3 ± 0.0	64.2 ± 0.6	0.42 ± 0.01	226 ± 4	173 ± 3
	13	0.04	2.6	8.30	7.4	2.1 ± 0.2	6.5 ± 0.7	0.0	134 ± 14	64 ± 7
	14	0.04	1.8	8.27	8.1	1.7 ± 0.0	7.6 ± 0.0	0.35	458 ± 9	269 ± 5
	25	0.04	1.5	8.28	7.9	1.3	20.5 ± 5	0.83 ± 0.07	ND	ND

*μl of cell volume per litre of culture.

†mmol of cellular DMSP, chlorophyll *a* (chl *a*), or total DMS per litre of cell volume. Total DMS was measured in unfiltered culture samples.

‡mmol ascorbate consumed per minute per litre cell volume (CV) or mol ascorbate per mol chl *a* per minute, respectively.

§ND: not determined.

||Mean ± range of two replicates for chl *a*, DMSP and DMS. APX activities are given as the mean ± s.d. for 2–4 replicates (mean 3.2).

chlorophyll *a*²⁰, which leads to inefficient photosynthetic electron transport²⁰. This, in turn, should increase oxidative stress¹⁹, which if not alleviated, could lead to severe cell damage or death.

Similar effects of CO₂ and Fe limitation were observed with other species. In the coccolithophore *Emiliania huxleyi*, which has constitutively high DMSP, CO₂ limitation caused up to a 150% increase in cellular DMSP and large relative increases in the DMS to cell-volume ratio (Table 3). For the coastal diatom *Skeletonema costatum*, Fe limitation caused a 67% decrease in growth rate and a three-fold increase in the cell DMSP/C ratio (from 1.6 to 4.7 mmol mol⁻¹) (W.S., D.J.K., R.P.K. and S.H., unpublished data). Fe limitation also has been found to increase cellular DMSP concentrations in *Phaeocystis* sp.¹¹.

The response of *E. huxleyi* to other oxidative stressors was also determined. Initial experiments examined the effect of solar UV radiation, an important environmental stress⁷. Growth of *E. huxleyi* cultures under 30% solar radiation, with different portions of the UV spectrum removed with cut-off filters, increased DMSP/cell-volume ratios by 5–100% and DMS/cell volume ratios by 2.4- to 36-fold relative to values in a culture grown under a similar intensity of fluorescent light with essentially no UV radiation (Table 4). The largest effect occurred in sunlight attenuated with Mylar or polycarbonate, which remove UV-B (290–320 nm) radiation but allow most UV-A (320–400 nm) to pass. Smaller increases occurred under full spectral sunlight or sunlight attenuated with UV-adsorbing glass, which removes >90% of the UV at wavelengths <383 nm. Exposure to UV oxidative stress should both promote metabolic induction of antioxidant systems⁷ and consume cellular antioxidants (DMSP and DMS). In addition, photolysis by UV in sea water can degrade DMS released from cells²¹. Consequently, the DMS measured under high levels of UV exposure may substantially underestimate overall DMS production. Because of these effects, the highest DMS accumulation should occur at intermediate levels of UV stress, as observed here in the Mylar and polycarbonate treatments.

In further support of our antioxidant hypothesis, the exposure of *E. huxleyi* to two additional oxidative stressors (high Cu²⁺ and H₂O₂) increased DMS/cell-volume ratios 9 to 15-fold (see Supplementary Information). Copper catalyses the reduction of H₂O₂ to ·OH radicals, and is a known oxidative stressor in phytoplankton⁹.

Our results showing increased ratios of DMS and DMSP to chlorophyll *a* with exposure to solar UV radiation (Table 4) are consistent with field data. Seasonally, maximum DMS concentrations in oceanic surface waters occur in the summer when solar UV exposure is highest owing to higher UV intensity, longer photoperiods, and shallower mixed layer depths^{22,23}. Surface DMS/chlorophyll *a* ratios increase by 10–50-fold from winter to summer in a broad range of subtropical, temperate and sub-arctic waters. This seasonal pattern has been suggested to be related to UV

photo-inhibition of bacteria which use DMS as a growth substrate²² and fluctuations in DMS photolysis rates²³; however, our results indicate that this pattern may also be due to higher algal DMSP production and increased lysis to DMS with higher solar UV stress. Much of the seasonal pattern may be related to the selection of high-DMSP species during the summer, capable of growth under high exposure to solar irradiance, such as *E. huxleyi*²⁴. But we argue that the high DMSP levels may help these species to adapt to the high summertime oxidative stress. Ratios of DMS, DMSP and DMSO to chlorophyll *a* are usually highest near the sea surface and decrease substantially with depth in conjunction with solar attenuation^{25,26}. Furthermore, in the euphotic zone of the tropical Atlantic, particulate DMSP (size range 0.7–10 μm) was closely correlated with levels of antioxidant carotenoids²⁶, again suggesting a direct linkage between oxidative stress and cellular levels of DMSP.

Linkages between solar UV radiation and algal DMSP and DMS production have important implications for large-scale feedback interactions between marine phytoplankton and atmospheric chemistry and physics^{1–3}. In these interactions, DMS diffuses into the atmosphere where it is oxidized to acidic species (for example, H₂SO₄) that form aerosols which serve as cloud condensation nuclei. The resultant cloud droplets increase solar reflectance, which has a cooling effect on the planet^{1,2}. There is an inherent negative feedback at work, because increased solar irradiance is associated with higher DMS concentrations at the ocean's surface^{1,22}, which our experiments suggest is at least partly linked to increased photo-oxidative stress. The resultant higher DMS flux to the atmosphere should increase the formation of acidic sulphur aerosols and clouds, thereby reducing solar inputs to the ocean's surface, which should have a cooling effect^{1,2} and reduce biological solar stress. Our results suggest that solar irradiance and other oxidative stressors, such as Fe limitation or CO₂ limitation, may act

Table 3 Effect of CO₂ limitation in cultures of *E. huxleyi*

Treatment	Day*	Growth rate (d ⁻¹)	Cell vol. (μl l ⁻¹)	pH	[CO ₂] (μmol l ⁻¹)	DMS/CV (mmol l ⁻¹)†	Cell DMSP (mmol l ⁻¹)‡
Control	10	0.73	8.1	8.46	4.6	0.8 ± 0.0	185 ± 0
	11	0.73	4.6	8.40	5.5	0.3 ± 0.0	167 ± 3
	12	0.73	4.9	8.41	5.3	0.1 ± 0.0	200 ± 10
	13	0.73	5.9	8.46	4.6	0.0 ± 0.0	193 ± 6
CO ₂ -limited	11	0.23‡	21.9	8.97	0.77	0.8 ± 0.0	232 ± 8
	12	0.23	21.5	8.92	0.93	1.1 ± 0.1	267 ± 6
	14	0.23	26.0	9.07	0.52	2.0 ± 0.2	302 ± 21
	17	0.09‡	29.6	9.26	0.25	4.7 ± 0.2	365 ± 10
	18	0.09	29.9	9.20	0.31	6.3 ± 0.1	396 ± 11
	19	0.09	29.4	9.18	0.34	5.5 ± 0.2	478
	20	0.09	30.3	9.14	0.40	5.3 ± 0.3	461 ± 17

*Days after culture inoculation.

†mmol total culture DMS or cellular DMSP per litre of cell volume. Data are given as the mean ± range for two replicates.

‡CO₂ limitation of growth rate was first evident on day 11. 0.23 d⁻¹ is the average specific growth rate for days 11–14 while 0.09 d⁻¹ is the average rate for days 14–20.

Table 4 Effect of solar radiation on DMSP and DMS in cultures of *E. huxleyi*

Treatment	Growth rate (d ⁻¹)	DMS/CV (mmol l ⁻¹) ^a	DMSP/CV (mmol l ⁻¹) ^a	chl a/CV (mmol l ⁻¹)	DMSP/chl a (mol mol ⁻¹)
Fluorescent light†	0.66	1.0 ± 0.0 (2)‡	198 ± 7 (2)	2.8	70
Exp. 1; 30% sun, UV A + B filter	0.64	4.7 ± 0.1 (2)	208 ± 5 (2)	3.2	65
Exp. 1; 30% sun, Mylar (UV-B filter)	0.57	36.2 ± 1.3 (2)	393 ± 6 (2)	3.9	101
Exp. 1; 30% sun, Full UV	0.65	2.4 ± 0.1 (2)	315 ± 1 (2)	3.6	86
Exp. 2; 30% sun, polycarbonate§	0.68	16.3 ± 0.9 (3)	274 ± 17 (4)	2.75 ± 0.02 (2)	96
Exp. 2; 10% sun, polycarbonate§	0.68	6.8 ± 0.6 (4)	256 ± 4 (4)	4.05 ± 0.08 (2)	63

^ammol of total culture DMS or DMSP normalized to cell volume (CV). All indoor and outdoor measurements were made 5–6 h into a 14-h light period.

†14 h d⁻¹ of fluorescent lighting at a PAR intensity of 800 μmol m⁻² s⁻¹.

‡The number of replicate measurements is given in parenthesis. The ± value denotes the range for *n* = 2 and s.d. for *n* > 2.

§Absorbs UV-B and >50% of UV-A at wavelengths <350 nm.

in concert as controls on DMS release from the ocean's surface, with important feedbacks on both climate and the productivity and composition of phytoplankton communities. □

Methods

Culture experiments

The CO₂ and Fe limitation experiments with *T. pseudonana* (clone CCMP1335) and *E. huxleyi* (CCMP374) were conducted at 20 °C under fluorescent light (Vita-Lite Plus, Durotest; intensity of 500 μE m⁻² s⁻¹). *T. pseudonana* and *E. huxleyi* were grown under light:dark cycles of 12 h:12 h and 14 h:10 h, respectively. Axenic cultures were grown in tightly sealed polycarbonate bottles with <5% headspace to minimize CO₂ and DMS exchange with the atmosphere. The culture media consisted of filtered sea water²⁷ enriched with 160 μM NaNO₃, 10 μM Na₂HPO₄, 40 μM Na₂SiO₃, 10 nM Na₂SeO₃, and vitamins (0.37 nM B₁₂, 2 nM biotin and 300 nM thiamin) and a 0.1 mM EDTA-trace-metal ion buffer system²⁷. The nitrate, phosphate and vitamins were five times normal levels²⁷ to ensure growth limitation by CO₂ at high biomass rather than by N and P. For the control and CO₂ limitation treatments, a non-limiting Fe level (1 μM)²⁷ was added.

The cells were grown semi-continuously by harvesting a portion of the culture periodically (usually every 1–3 d) and replacing it with fresh medium. For the control, cell volumes were maintained at ≤8 μl per litre of culture to avoid nutrient limitation. To achieve CO₂ limitation, cultures were grown to high cell volumes, which resulted in up to a 40-fold decrease in CO₂ concentrations (computed from the CO₂ alkalinity, pH and conditional constants for CO₂ and bicarbonate dissociation). In the Fe-limited *T. pseudonana* treatment, cells were transferred from an Fe-sufficient medium (1 μM Fe) to a medium with no added Fe and grown into Fe limitation. The harvested portions of all cultures were analysed for total cell volume²⁷, pH, cellular chlorophyll *a*²⁷, total DMS²⁸ and particulate (that is, cellular) DMSP²⁸. In addition, APX activity was measured¹⁸ in *T. pseudonana*. For control and Fe-limited cultures, the specific growth rate was computed by linear regression of ln cell volume versus time²⁷ after accounting for culture dilution effects.

The two solar exposure experiments (Table 4) were conducted with axenic cultures of *E. huxleyi* (CCMP374) grown in seawater media similar to the above, but with five-fold lower NaNO₃ (32 μM), Na₂HPO₄ (2 μM) and vitamins. In addition, the metal ion buffering agent was switched from 0.1 mM EDTA to 0.1 mM NTA (see Supplementary Information) to avoid photolysis of FeEDTA²⁷. The experiments were initiated by growing a non-limited control culture and two experimental cultures under 14 h d⁻¹ of fluorescent light (Cool White, Philips). The light intensity was 800 μE m⁻² s⁻¹, approximating that of 30% solar radiation (see below). The experimental cultures were transferred to media held in 80-ml sealed quartz tubes (exp. 1) or in 150-ml polycarbonate plastic bottles (exp. 2) all with minimal headspace. The cultures were exposed to sunlight attenuated to 30% (exp. 1) and to 30% and 10% (exp. 2) using neutral density screens. In addition, two of the three quartz tubes in exp. 1 were covered with a UV cut-off filter: Mylar, which absorbed >90% of solar UV radiation at wavelengths <317 nm, or UV-absorbing glass, which absorbed >90% of radiation at wavelengths <383 nm. The polycarbonate bottles constitute a short-wavelength UV filter as they absorb >50% of UV radiation at wavelengths <350 nm, and >90% at wavelengths <330 nm. In exps 1 and 2 the cultures experienced three- and five-day growth lags, respectively, while acclimating to the new light conditions, and then grew exponentially. Total DMS²⁸, total DMSP²⁸, cell volume²⁷ and cell chlorophyll *a*²⁷ were measured in the cultures after three days of exponential growth in exp. 1 and five days in exp. 2. The same measurements were made after 10 d of exponential growth in the control. Specific growth rates in these cultures were measured²⁷ during the exponential growth period. The two solar experiments were conducted at ~23 °C in a flowing water bath under mostly sunny skies during May 2001 (latitude 35° N). Sterile technique was used in all experiments (Tables 2–4) to avoid bacterial contamination (see Supplementary Information).

In early tests, the measured cell DMSP concentration decreased with increasing volume filtered (1–10 ml) in *E. huxleyi* under gentle vacuum (<5 mm Hg), indicating some loss of DMSP during filtration due to cell lysis. This problem did not occur with *T. pseudonana*, which has a robust silica frustule. Because of these problems with *E. huxleyi*, filtration was avoided and total culture DMSP was measured in early experiments with this species, including the solar exposure experiment (Table 4). Recently, we found that loss of cell DMSP can be largely eliminated by using gravity filtration and a small filtration volume (1–2 ml). This procedure was used to determine cellular DMSP in the CO₂-limitation experiment with *E. huxleyi* (Table 3). In this experiment, DMSP concentrations measured in the culture filtrates (data not shown) were 2.3 ± 0.7% (*n* = 8) and 2.6 ± 1.2%

(*n* = 12) of the total culture DMSP for the control and CO₂-limited treatments. Thus, virtually all of the DMSP is present in the cells. Furthermore, the large variability in DMSP in the filtrates (up to 2.8-fold between duplicate filtrates) may be due to a variable residual lysis of cells under even these gentlest of filtration conditions.

Hydroxyl radical reaction with DMSP, DMS, DMSO and MSNA

Laboratory experiments at 25 °C and pH 7.5 confirmed that •OH radicals oxidize DMS to DMSO, MSNA and methane sulphonate (MSA), and quantitatively oxidize DMSO stepwise to MSNA and MSA. DMSP was likewise oxidized quantitatively to MSNA and MSA, with DMSO as a likely reactive intermediate. In these experiments, rate constants for reactions of DMSP and MSNA with the •OH radical were also measured by comparison with the reaction kinetics of DMSO under identical conditions. The rate constant of DMSO with •OH under these conditions was 6.6 × 10⁹ M s⁻¹ (ref. 5). To generate the •OH radical, a 10 mM aqueous solution of H₂O₂ in 10 mM phosphate buffer (pH 7.5) was irradiated under UV lamps (UV-A, Q-Panel) with UV-A and UV-B irradiances of 3.0 and 0.28 μW cm⁻². Loss of DMSP and DMSO were measured over time with gas chromatography²⁸, and loss of MSNA and its associated conversion to MSA were measured by ion chromatography. Rate constant calculations were corrected for direct oxidation by H₂O₂ measured in parallel solutions in the dark. Nearly identical results were obtained when nitrate photolysis was used to generate •OH.

Received 7 February; accepted 2 May 2002; doi:10.1038/nature00851.

- Bates, T. S., Charlson, R. J. & Gammon, R. H. Evidence for the climatic role of marine biogenic sulphur. *Nature* **329**, 319–321 (1987).
- Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. Oceanic phytoplankton, atmospheric sulphur, cloud albedo, and climate. *Nature* **326**, 655–661 (1987).
- Andreae, M. O. & Crutzen, P. J. Atmospheric aerosols: Biogeochemical sources and role in atmospheric chemistry. *Science* **276**, 1052–1058 (1997).
- Stefels, J. Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. *J. Sea Res.* **43**, 183–197 (2000).
- Buxton, G. V., Greenstock, C. L., Helman, W. P. & Ross, A. B. Critical review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals in aqueous solution. *J. Phys. Chem. Ref. Data* **17**, 513–886 (1988).
- Jakob, B. & Heber, U. Photoproduction and detoxification of hydroxyl radicals in chloroplasts and leaves and (its) relation to photoinactivation of photosystems I and II. *Plant Cell Physiol.* **37**, 625–629 (1996).
- Lesser, M. P. & Shick, J. M. Effects of irradiance and ultraviolet radiation on photo-adaptation in the zooxanthellae of *Aiptasia pallida*: primary production, photoinhibition, and enzymatic defenses against oxygen toxicity. *Mar. Biol.* **102**, 243–255 (1989).
- Butow, B., Wynne, D., Sukenik, A., Hadas, O. & Tel-Or, E. Synergistic effect of carbon concentration and high temperature on lipid peroxidation in *Peridinium gatunense*. *J. Plankt. Res.* **20**, 355–369 (1998).
- Okamoto, O. K., Pinto, E., Latorre, L. R., Bechara, E. J. H. & Colepicolo, P. Antioxidant modulation in response to metal-induced oxidative stress in algal chloroplasts. *Arch. Environ. Contam. Toxicol.* **40**, 18–24 (2000).
- Karsten, U., Wiencke, C. & Kirst, G. O. The effect of light intensity and day length on the dimethylsulphoniopropionate (DMSP) content of green marine macroalgae from Antarctica. *Plant Cell Environ.* **13**, 989–993 (1990).
- Stefels, J. & van Leeuwen, M. A. Effects of iron and light stress on the biochemical composition of antarctic *Phaeocystis* sp. (Prymnesiophyceae). I. Intracellular DMSP concentrations. *J. Phycol.* **34**, 486–495 (1998).
- Steinke, M., Daniel, C. & Kirst, G. O. In *Biological and Environmental Chemistry of DMSP and Related Sulphonium Compounds* (eds Kiene, R. P., Visscher, P. T., Keller, M. D. & Kirst, G. O.) 317–324 (Plenum, New York, 1996).
- Wilkinson, F., Helman, W. P. & Ross, A. B. Rate constants for the decay and reactions of the lowest electronically excited singlet state of molecular oxygen in solution. An expanded and revised compilation. *J. Phys. Chem. Ref. Data* **24**, 663–1021 (1995).
- Simó, R., Hattori, A. D., Malin, G. & Liss, P. S. Particulate dimethylsulphoxide in seawater: production by microplankton. *Mar. Ecol. Prog. Ser.* **167**, 291–296 (1998).
- Lee, P. A. & de Mora, S. J. Intracellular dimethylsulfoxide (DMSO) in unicellular marine algae: Speculations on its origin and possible biological role. *J. Phycol.* **35**, 8–18 (1999).
- Simó, R., Pedrós-Allió, C., Malin, G. & Grimalt, J. O. Biological turnover of DMS, DMSP, and DMSO in contrasting open-sea waters. *Mar. Ecol. Prog. Ser.* **203**, 1–11 (2000).
- Scaduto, R. C. Oxidation of DMSO and methane sulfinic acid by the hydroxyl radical. *Free Radic. Biol. Med.* **18**, 271–277 (1995).
- Asada, K. & Nakano, Y. Hydrogen peroxide is scavenged by ascorbate-specific peroxidase in spinach chloroplasts. *Plant Cell Physiol.* **22**, 867–880 (1981).

19. Niogi, K. K. Photoprotection revisited: Genetic and molecular approaches. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 333–359 (1999).
20. Greene, R. M., Geider, R. J., Kolber, Z. & Falkowski, P. G. Iron-induced changes in light-harvesting and photochemical energy conversion processes in eukaryotic marine algae. *Plant Physiol.* **100**, 565–575 (1992).
21. Kieber, D. J., Jiao, J., Kiene, R. P. & Bates, T. S. The impact of dimethyl sulfide photochemistry on methyl sulfur cycling in the Equatorial Pacific Ocean. *J. Geophys. Res.* **101**, 3715–3722 (1996).
22. Simó, R. & Pedrós-Allió, C. Role of vertical mixing in controlling the oceanic production of dimethyl sulphide. *Nature* **402**, 396–399 (1999).
23. Dacey, J. W., Howse, F. A., Michaels, A. F. & Wakeham, S. G. Temporal variability of dimethylsulfide and dimethylsulfoniopropionate in the Sargasso Sea. *Deep-Sea Res.* **1** **45**, 2085–2104 (1998).
24. Naninga, H. J. & Tyrrell, T. Importance of light for the formation of algal blooms by *Emiliania huxleyi*. *Mar. Ecol. Prog. Ser.* **136**, 195–206 (1996).
25. Simó, R., Grimalt, J. O., Pedrós-Alio, C. & Albaiges, J. Occurrence and transformation of dissolved dimethyl sulfur species in stratified seawater (western Mediterranean Sea). *Mar. Ecol. Prog. Ser.* **127**, 291–299 (1995).
26. Belviso, S. *et al.* Size distribution of dimethylsulfoniopropionate (DMSP) in areas of the tropical northeastern Atlantic Ocean and the Mediterranean Sea. *Mar. Chem.* **44**, 55–71 (1993).
27. Sunda, W. G. & Huntsman, S. A. Iron uptake and growth limitation in oceanic and coastal phytoplankton. *Mar. Chem.* **50**, 189–206 (1995).
28. Kiene, R. P. & Gerard, G. Determination of trace levels of dimethyl sulfoxide (DMSO) in seawater and rainwater. *Mar. Chem.* **47**, 1–12 (1994).
29. Ahner, B. A., Wei, L., Oleson, J. R. & Ogura, N. Glutathione and other low molecular weight thiols in marine phytoplankton under metal stress. *Mar. Ecol. Prog. Ser.* **232**, 93–103 (2002).
30. Dabrowski, K. *Ascorbate in Aquatic Organisms* (CRC, Boca Raton, 2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank D. Blackwood, P. Griffin and L. Linn for technical assistance.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.S. (e-mail: bill.sunda@noaa.gov).

Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*

John C. Wootton*, Xiaorong Feng†, Michael T. Ferdig‡, Roland A. Cooper†, Jianbing Mu†, Dror I. Baruch†, Alan J. Magill†§ & Xin-zhuan Su†

* Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894-6075, USA

† Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892-0425, USA

‡ Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556-016, USA

§ Division of Communicable Disease and Immunology, Walter Reed Army Institute of Research, Washington DC 20307-5100, USA

Widespread use of antimalarial agents can profoundly influence the evolution of the human malaria parasite *Plasmodium falciparum*. Recent selective sweeps for drug-resistant genotypes may have restricted the genetic diversity of this parasite, resembling effects attributed in current debates^{1–4} to a historic population bottleneck. Chloroquine-resistant (CQR) parasites were initially reported about 45 years ago from two foci in southeast Asia and South America⁵, but the number of CQR founder mutations and the impact of chloroquine on parasite genomes worldwide have been difficult to evaluate. Using 342 highly polymorphic microsatellite markers from a genetic map⁶, here we show that the level

of genetic diversity varies substantially among different regions of the parasite genome, revealing extensive linkage disequilibrium surrounding the key CQR gene *pfcr7* and at least four CQR founder events. This disequilibrium and its decay rate in the *pfcr7*-flanking region are consistent with strong directional selective sweeps occurring over only ~20–80 sexual generations, especially a single resistant *pfcr7* haplotype spreading to very high frequencies throughout most of Asia and Africa. The presence of linkage disequilibrium provides a basis for mapping genes under drug selection in *P. falciparum*.

We examined microsatellite markers covering the 14 haploid chromosomes, at an average interval of ~75 kilobases (kb), from 87 *P. falciparum* worldwide isolates (Supplementary Information Fig. 1). Eighty-five per cent of the markers are highly polymorphic with 10 or more alleles. To evaluate whether genotypes reflect the geographical origins of the isolates, we tested the hypothesis of no allele sharing within a continental subset of isolates (Fig. 1a). The American isolates show very significant allele sharing for all chromosomes ($P = 1.5 \times 10^{-37}$ genome-wide), as do Asian isolates for most chromosomes except 1, 2, 8 and 14 ($P = 2.15 \times 10^{-7}$). In contrast, African isolates are very diverse and have a level of allele sharing similar to the bootstrap controls. This analysis adds genome-wide support to previous claims^{8–10} that distinct parasite population structures exist on different continents, and that African parasites could be the common ancestors of *P. falciparum* parasites worldwide because of their higher diversity.

However, substantial differences in allelic diversity exist among chromosomes (Fig. 1a), possibly reflecting recent directional selection superimposed on the patterns of geographical diversity. The CQR isolates from Africa and Asia shared many more alleles on chromosome 7 (where *pfcr7* is located) than the chloroquine sensitive (CQS) isolates (Fig. 1b), suggesting common origins for CQR parasites. To further explore the extent of allele sharing on chromosome 7, we analysed the *pfcr7* alleles and flanking

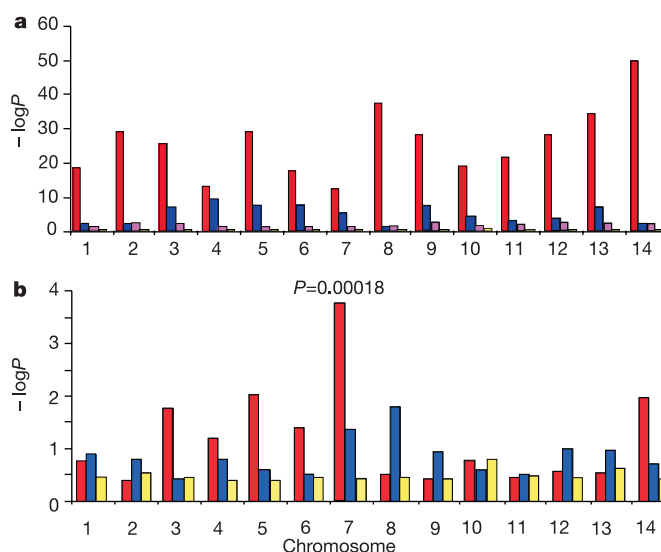


Figure 1 Genome-wide allele sharing analysis of *P. falciparum* isolates from various geographical regions or CQR/CQS (chloroquine-resistant/chloroquine sensitive) subsets. For the individual chromosomes, bars represent the negative $\log_{10}(P)$ values for obtaining by chance the amount of allele sharing observed, based on 20,000 bootstrap samples (see Methods; higher bars indicate more significant allele sharing). **a**, Allele sharing among parasites from Asia (blue), Africa (purple) and South America (red). Yellow, bootstrap controls. **b**, Higher allele sharing among CQR (red) than CQS (blue) parasites from Africa and Asia for chromosome 7 ($P = 1.76 \times 10^{-4}$) and to a lesser extent for chromosomes 3, 5 and 14. Yellow, bootstrap controls.

microsatellite haplotypes.

We found that nearly all CQR isolates from both Papua New Guinea and the Brazilian/Peruvian Amazon have the same *pfcr* allele, coding for five amino-acid substitutions (S-MNT-H-S-Q-D-L-R, Fig. 2a). This allele probably evolved independently in South America and Papua New Guinea because the *pfcr*-flanking microsatellite haplotypes are notably different (Fig. 2a); and the genome-wide allele sharing between Papua New Guinea and the Amazon is not significantly different from the background sharing between other Asian and American isolates ($P = 0.34$). In contrast, a distinct allele (C-MET-Q-S-Q-N-I-T) in Columbian parasite JAV evolved with an ECU-like flanking haplotype and genetic background (see Fig. 2a). African and Asian (excluding Papua New Guinea) CQR parasites have a near-identical *P. falciparum* chloroquine-resistance transporter (PfCRT) with 6–8 amino-acid substitutions and a common *pfcr*-flanking haplotype (Fig. 2a). Asian isolates (except FCB and its close relatives) have a C-IET-H-S-E-S-T-I variant,

whereas most African isolates and FCB have C-IET-H-S-E-S-I-I, suggesting that a back mutation in an Asian CQR parasite pre-dated the spread of CQR parasites in Africa. These *pfcr* alleles, the flanking microsatellite markers, and the lack of admixture shown by genome-wide haplotypes demonstrate—in contrast to the two origins previously proposed⁵—at least four independent CQR founder events: one in Asia spreading to Africa, one in Papua New Guinea, and two in South America.

A hallmark of a selective sweep is a chromosomal region with reduced allelic diversity associated with a specific phenotype. One example is the *pfcr* allele (C-IET-H-S-E-S-T-I) that has overwhelmed most of Asia and Africa with current allele frequencies of ~40–100% (refs 11, 12). Chromosomal segments spanning >200 kb surrounding *pfcr* have been maintained in most of the CQR parasites, but decline gradually with distance, whereas those in the CQS parasites are highly diverse in these flanking regions (Fig. 2a–c). On the basis of the high recombination rates inferred

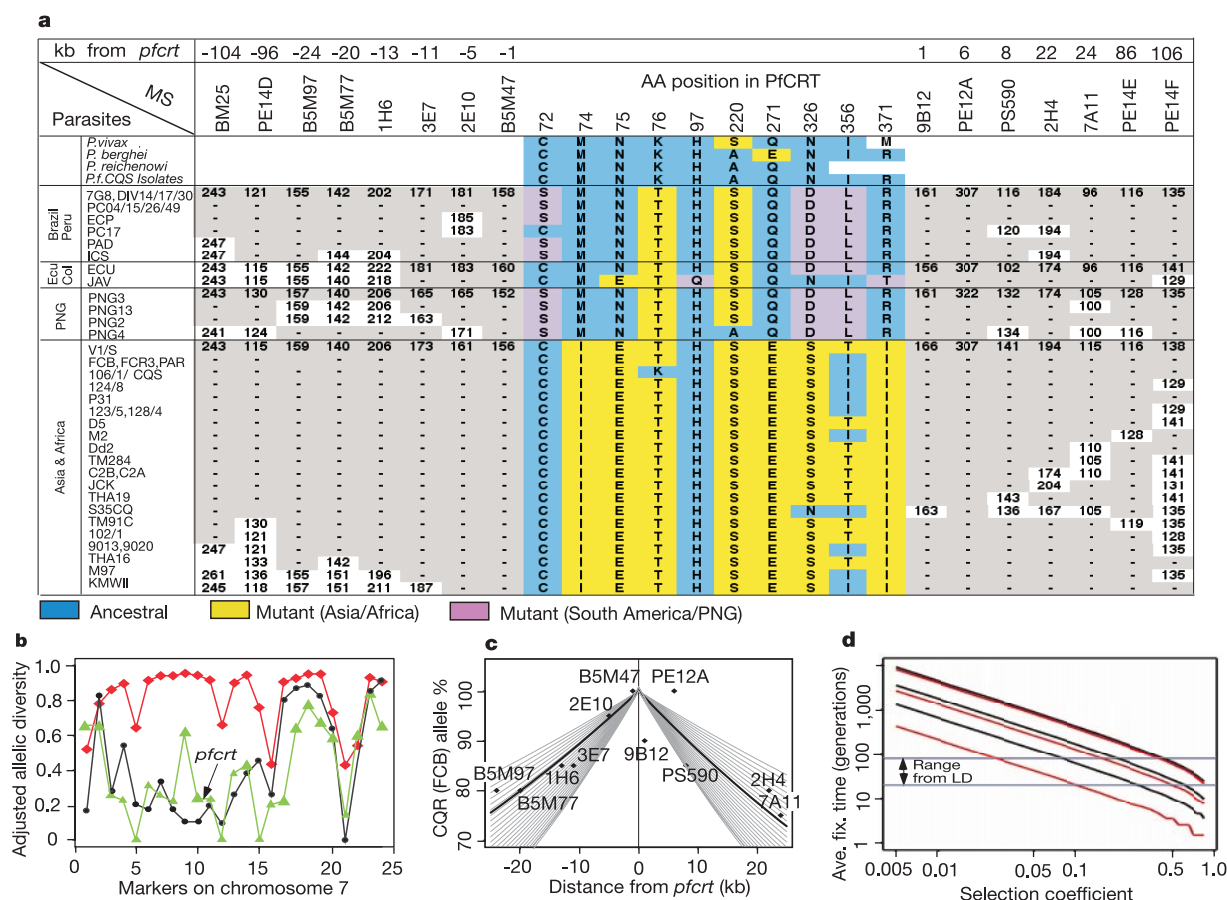


Figure 2 Extensive linkage disequilibrium (LD), microsatellite (MS) haplotypes flanking *pfcr*, and the rate of LD decline in CQR isolates. **a**, Amino-acid substitutions in PfCRT and its flanking MS haplotypes. MS and their distances from *pfcr* are indicated at the top. 'Pf CQS isolates' represents 38 CQS isolates having the ancestral C-MNK-H-A-Q-N-I-R allele (blue shading), but with highly variable flanking MS haplotypes (Supplementary Information Table 1). CQR parasites are grouped according to their genotypes and geographic origins (Ecu, Ecuador; Col, Columbia); and dashes represent the same MS allele as the one at the top of each section. The substitutions in some *P. falciparum* isolates and the orthologues from *Plasmodium vivax*, *Plasmodium berghei*, and *Plasmodium knowlesi* are from refs 7 and 20. **b**, Decreased diversity in CQR isolates surrounding *pfcr* (arrow) suggests chloroquine selection. Black, CQR parasites from Asia and Africa; green, CQR parasites from South America; red, CQS parasites from Asia and

Africa. **c**, Decline of LD from *pfcr* (at position 0) in Asian/African CQR isolates. The points represent the percentage of shared alleles of each marker. The thin lines are the expected LD decay curves obtained by simulation from equation (2), using known recombination rate⁶, an inbreeding coefficient of 0.5 and a range, in steps of 2, of 25 (top line) to 75 sexual generations. The thick lines are the best-fit regression curves corresponding to ~40 generations for inbreeding coefficient of 0.5. **d**, Estimation of the selection coefficients and numbers of generations required to fix a drug resistant allele under different initial allele proportions (p_0): 10^{-18} (upper lines); 10^{-6} (middle lines); 10^{-1} (lower lines). Final allele proportions (p_f): 0.5 (red lines); 0.99 (black lines). The results show no possibility of generating the observed high CQR *pfcr* allele frequencies by genetic drift in the 20–80 generation period (horizontal lines) inferred from LD.

from a genetic cross⁶ and field populations^{8,13}, and inbreeding coefficients previously estimated for Africa and Asia as 0.3–0.9 (refs 14–16), we estimate that this linkage disequilibrium decay has occurred over 20–80 parasite sexual generations. The timescale corresponding to 20–80 generations is ~6–30 years in endemic areas, emphasizing the profound influence of chloroquine on recent *P. falciparum* evolution.

Fixation of this CQR haplotype by genetic drift alone would require hundreds or thousands of generations (Fig. 2d). Although genetic drift, especially during epidemic expansions, could fix neutral alleles in isolated populations, neutral mechanisms cannot fix the same allele as observed for CQR on the continental scale across many parasite populations with different endemicity and transmission levels. Simulations using a wide range of initial allele frequencies to model the effects of different rates of migration and mutation showed that very high selection coefficients of about 0.1–0.7 are necessary to produce the level of linkage disequilibrium observed (Fig. 2c, d, and Supplementary Information Fig. 2). These are substantially larger than selection coefficients (<0.0002) estimated from effective population sizes^{3,8} that would typically fix effectively-neutral alleles by stochastic dynamics. From computer simulation (Supplementary Information Fig. 2), linkage disequilibrium segment lengths encompassing the *pfcr* locus after 40 generations of directional selection are consistent with the 10–200 kb observed.

Only a small repertoire of amino-acid substitutions in PfCRT appears to be compatible with the constraints imposed by drug selection and the biological function of the protein. A minimum of three changes, at positions that are highly conserved in different *Plasmodium* species and CQS parasites, are present in addition to the critical 76T substitution (Fig. 2a). These substitutions may have evolved to compensate deleterious effects of 76T and/or as quantitative modulators of the chloroquine response. To investigate if compensatory mutations elsewhere in the genome have evolved under chloroquine selection together with *pfcr* mutations, we

mapped potential regions of linkage disequilibrium using genome-wide scans (Fig. 3). We found evidence of reduced allelic diversity, additional to *pfcr*, especially on chromosomes 1, 5, 6, 7, 10 and 12 (Fig. 3a, b). However, none of these loci (except the *pfcr* locus) can be attributed directly to the chloroquine selective sweep because most of the peaks are present in both CQS and CQR isolates. The flat baselines (Fig. 3c) in African isolates suggest that no other genes with effects similar to *pfcr* spread throughout the continent, although genes with minor effects could escape our detection at present marker resolution.

Our genome-wide analysis shows highly diverse *P. falciparum* populations, with variable levels of polymorphisms in different chromosomal regions. Such a genomic structure is consistent with a history of multiple selective sweeps, and raises questions concerning the debate on a recent major bottleneck of *P. falciparum* evolution^{1–4}. Successive selective sweeps or a bottleneck could present a picture of limited genetic diversity if relatively few loci are tested in local populations. Concerted genome-wide analyses of both single nucleotide polymorphism and microsatellite variation may give further insight into these and other questions related to diversifying, directional, stabilizing and balancing selection in *P. falciparum* evolution.

The reduced diversity and shared chromosomal segments at the *pfcr* locus show that linkage-disequilibrium based methods could be used to map *P. falciparum* genes under drug selection. Association studies with other antimalarial drugs, singly or in combination, could reveal new drug-response genes and contribute further information on the diversity and evolutionary potential of the *P. falciparum* genome. □

Methods

Genotyping, DNA sequencing and drug assay

Parasite culture and DNA extraction were as described^{17,18}. For microsatellite typing, polymerase chain reaction (PCR) products were labelled with fluorescent dyes and detected on an ABI377 automatic sequencer as described⁶, using ROX 500XL size standards (PE Biosystem) and Genotyper software. PCR products amplified from genomic DNA were sequenced directly without cloning. Half maximal inhibitory concentrations (IC₅₀s) of the isolates in responses to chloroquine were determined as described¹⁸ after culturing the parasites for 48 h followed by the addition of ³H hypoxanthine for another 24 h.

Statistical and computational analysis

To analyse allele sharing in a geographical or phenotypic subset of *n* isolates (Fig. 1), an average haplotype distance was measured:

$$D_{av} = \left(\frac{1}{n^2} \right) \sum_{i=1}^n \sum_{j=1}^n (M_{T,ij} / M_{S,ij}) \quad (1)$$

M_T is the total number of markers compared between isolate *i* and *j*, and M_S is the number of shared alleles. Negative log probabilities of D_{av} were estimated using empirical distributions obtained from 20,000 bootstrap samples of randomly picked isolates.

Decline of linkage disequilibrium was calculated using¹⁹

$$L_i = 100[1 - d_i c(1 - F)]^g \quad (2)$$

where d_i is the distance (kb) between the gene under selection (*pfcr*) and the *i*th marker, L_i is the percentage of isolates in which the entire progenitor haplotype was maintained over d_i , c is the crossover frequency per meiosis (one crossover per 1,700 kb per meiosis⁶), F is the inbreeding coefficient, and g is the number of sexual generations.

Changes in frequencies of two alleles (*R* and *S*) conferring drug resistance versus sensitivity were modelled for haploid inheritance using:

$$\ln \left(\frac{p_g}{1 - p_g} \right) = \ln \left(\frac{p_0}{1 - p_0} \right) + g \ln \left(\frac{1}{1 - s} \right) \quad (3)$$

where p_0 and p_g are the proportions of the *R* allele, respectively, initially and at generation *g*, and *s* is the selection coefficient against the *S* allele (relative fitness of *R* and *S*: 1 and $(1 - s)$ respectively). To encompass the effects of a wide range of population structures and initial conditions, p_0 was varied from 0.1 (high introduction frequency of the *R* allele by human migration) to 10^{-24} (the estimated frequency of a double or triple mutation arising *de novo*). Values of p_g were computed over 10,000 generations for *s* values ranging from 0.0005 (near-neutrality) to 1, to determine the average times required to give *R* allele p_g values of 0.5 to 0.99.

The 'adjusted allelic diversity' (Fig. 3a, b) is heterozygosity normalized for the marker-specific stepwise mutational component. The estimated number (*A*) of 2-bp or 3-bp steps available to a marker was computed from the range and intervals observed in its allele size

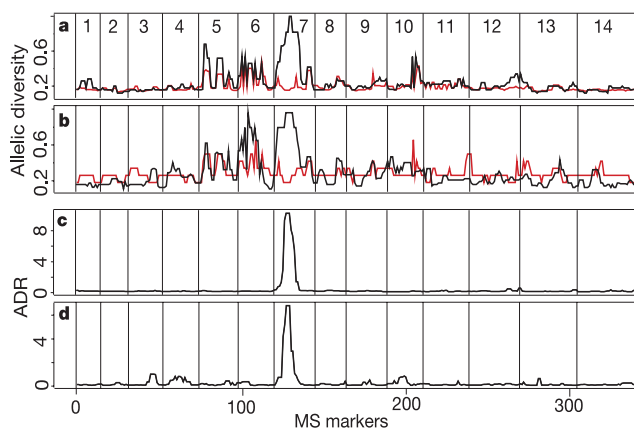


Figure 3 Genome-wide scans for loci of reduced diversity and association of reduced diversity with the CQR phenotype. The numbered panels represent the 14 chromosomes covered by 342 markers. **a, b**, Allelic diversity, plotted as $(1 - \text{median adjusted allelic diversity of five contiguous markers})$, for CQS (red) and CQR (black) isolates from Africa (**a**) and Asia (**b**, excluding CQR PNG isolates). Peaks represent regions with reduced diversity. **c, d**, allelic diversity ratio (ADR, plotted as the median log likelihood statistic over three contiguous markers) comparing CQR and CQS from Africa (**c**) and Asia (**d**), respectively. In both cases, a highly significant peak of log likelihood ratio at the *pfcr* locus on chromosome 7 was correctly identified, demonstrating the power of this approach for detecting drug-resistant genes in malaria parasites. Peaks with $\text{ADR} < 3$ are not statistically significant.

distribution. Then:

$$Y = \frac{A}{A+1} \left[1 - \sum_i p_i^2 \right] \quad (4)$$

where p_i is the proportion of the i th allele.

The allelic diversity ratio (ADR, Fig. 3c, d) was used to detect significant diversity differences between two phenotypically distinct sets of isolates (CQS and CQR):

$$\text{ADR} = \frac{\left(\frac{n_1}{n_1+1} \right) \sum_{i=1}^{k_1} p_i^2}{\left(\frac{n_2}{n_2+1} \right) \sum_{j=1}^{k_2} p_j^2} \quad (5)$$

where the isolates of the two phenotypes have n_1 and n_2 individuals and k_1 and k_2 alleles at proportions p_i and p_j . Log likelihoods for ADR were obtained empirically using 1,000 permutations of the indices of all isolates over their fixed genome-wide microsatellite haplotypes together with an empirical correction for the degree of relatedness among isolates. ADR, which cancels out biases owing to different intrinsic microsatellite mutation modes and homoplasy rates, was determined for all individual markers, then plotted (Fig. 3c, d) as the median filtered log likelihood over three-marker windows.

Received 26 September 2001; accepted 28 March 2002; doi:10.1038/nature00813.

- Rich, S. M. & Ayala, F. J. The recent origin of allelic variation in antigenic determinants of *Plasmodium falciparum*. *Genetics* **150**, 515–517 (1998).
- Volkman, S. K. *et al.* Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* **293**, 482–484 (2001).
- Hughes, A. L. & Verra, F. Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proc. R. Soc. Lond. B* **268**, 1855–1860 (2001).
- Hey, J. Parasite populations: the puzzle of *Plasmodium*. *Curr. Biol.* **9**, R565–R567 (1999).
- Payne, D. Spread of chloroquine resistance in *Plasmodium falciparum*. *Parasitol. Today* **3**, 241–246 (1987).
- Su, X. *et al.* A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* **286**, 1351–1353 (1999).
- Fidock, D. A. *et al.* Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell* **6**, 861–871 (2000).
- Anderson, T. J. *et al.* Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17**, 1467–1482 (2000).
- Escalante, A. A., Barrio, E. & Ayala, F. J. Evolutionary origin of human and primate malaria: evidence from the circumsporozoite protein gene. *Mol. Biol. Evol.* **12**, 616–626 (1995).
- Conway, D. J. *et al.* Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol. Biochem. Parasitol.* **111**, 163–171 (2000).
- Djimde, A. *et al.* A molecular marker for chloroquine-resistant *falciparum* malaria. *N. Engl. J. Med.* **344**, 257–263 (2001).
- Dorsey, G., Kanya, M. R., Singh, A. & Rosenthal, P. J. Polymorphisms in the *Plasmodium falciparum* pfcr and pfmdr-1 genes and clinical response to chloroquine in Kampala, Uganda. *J. Infect. Dis.* **183**, 1417–1420 (2001).
- Conway, D. J. *et al.* High recombination rate in natural populations of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **96**, 4506–4511 (1999).
- Hill, W. G., Babiker, H. A., Ranford-Cartwright, L. C. & Walliker, D. Estimation of inbreeding coefficients from genotypic data on multiple alleles, and application to estimation of clonality in malaria parasites. *Genet. Res.* **65**, 53–61 (1995).
- Walliker, D., Babiker, H. & Ranford Cartwright, L. in *Malaria: Parasite Biology, Pathogenesis, and Protection* (ed. Sherman, I. W.) 235–252 (American Society for Microbiology, Washington DC, 1998).
- Paul, R. E. *et al.* Mating patterns in malaria parasite populations of Papua New Guinea. *Science* **269**, 1709–1711 (1995).
- Trager, W. & Jensen, J. B. Human malaria parasites in continuous culture. *Science* **193**, 673–675 (1976).
- Su, X., Kirkman, L. A., Fujioka, H. & Wellems, T. E. Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant *P. falciparum* in Southeast Asia and Africa. *Cell* **91**, 593–603 (1997).
- Dye, C. & Williams, B. G. Multigenic drug resistance among inbred malaria parasites. *Proc. R. Soc. Lond. B* **264**, 61–67 (1997).
- Nomura, T. *et al.* Evidence for different mechanisms of chloroquine resistance in 2 *Plasmodium* species that cause human malaria. *J. Infect. Dis.* **183**, 1553–1561 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank various investigators who provided the isolates over the years, S. Davis-Hayman, D. Joy, K. Hayton and B. Marshall for critical reading of the manuscript and editorial assistance, and T. Wellems, L. Miller and D. Lipman for support and encouragement. The opinions of the authors do not necessarily reflect those of the US army or the Department of Defense.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.-z.S. (e-mail: xsu@niaid.nih.gov).

Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*

Jianbing Mu*, Junhui Duan*, Kateryna D. Makova†, Deirdre A. Joy*, Chuong Q. Huynh‡, Oralee H. Branch§, Wen-Hsiung Li† & Xin-zhuan Su*

* Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892-0425, USA

† Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA

‡ Information Engineering Branch, § Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894-6075, USA

The Malaria's Eve hypothesis, proposing a severe recent population bottleneck (about 3,000–5,000 years ago) of the human malaria parasite *Plasmodium falciparum*, has prompted a debate about the origin and evolution of the parasite^{1–6}. The hypothesis implies that the parasite population is relatively homogeneous, favouring malaria control measures. Other studies, however, suggested an ancient origin and large effective population size^{5,7–10}. To test the hypothesis, we analysed single nucleotide polymorphisms (SNPs) from 204 genes on chromosome 3 of *P. falciparum*. We have identified 403 polymorphic sites, including 238 SNPs and 165 microsatellites, from five parasite clones, establishing chromosome-wide haplotypes and a dense map with one polymorphic marker per ~2.3 kilobases. On the basis of synonymous SNPs and non-coding SNPs, we estimate the time to the most recent common ancestor to be ~100,000–180,000 years, significantly older than the proposed bottleneck. Our estimated divergence time coincides approximately with the start of human population expansion¹¹, and is consistent with a genetically complex organism able to evade host immunity and other antimalarial efforts.

The launching of the international project to sequence the genome of *P. falciparum*, and the completion of chromosomes 2 and 3 (refs 12, 13), have enabled us to search directly for SNPs from the parasite genes. Here we analyse coding sequences (202 kilobases, kb) and non-coding sequences (16 kb) on chromosome 3 from five independent parasite clones originating from Southeast Asia (Dd2), Africa (3D7, sequences obtained from PlasmoDB; see Methods), South America (7G8), Central America (Hb3) and Papua New Guinea (D10). Their identities and geographical origins were verified by genotyping the parasite DNA with 348 microsatellite (MS) markers, which showed highly diverse genetic backgrounds (mean genetic diversity $H = 0.8776 \pm 0.0079$). Oligonucleotide primers were synthesized to amplify 1.2–1.5 kb DNA fragments from each predicted gene. Polymerase chain reaction (PCR) products were sequenced directly without cloning into bacteria, and SNPs were identified after alignment of the DNA sequences. We also independently amplified and sequenced DNA covering 56 SNPs, including 32 randomly chosen and all 24 SNPs with single strand coverage, to validate the accuracy of the SNPs. Fifty-four of the SNPs were verified (~96%). Both of the two false SNPs were base call errors from regions of single strand coverage.

Despite not being distributed evenly, SNPs occur across the chromosome, with most of the genes (80.5%) containing only one or two SNPs (Fig. 1a, b). Of the 238 SNPs identified from 118 of 204 predicted genes (57.8%), including 153 complete or partial introns (Table 1), 207 were in coding regions (cSNP) and 31 in non-coding regions (ncSNP), although the frequency of SNPs was greater in non-coding regions (Table 1). There were more than

twice as many non-synonymous SNPs (nsSNPs) as synonymous SNPs (sSNPs) in the coding regions, in sharp contrast to similar frequencies of sSNP and nsSNP or slightly more sSNPs in the human genome^{14,15}. We also identified 165 polymorphic MS from 117 genes; 86 of the MS are in introns (Fig. 1a, b). This high frequency of MS is consistent with previous observations^{2,16} and implies that polymorphic MS can be as abundant as SNPs in this parasite. This collection of SNPs and MS, in conjunction with the 64 MS markers we developed previously¹⁶, constitutes a high-resolution map (Fig. 1a) of chromosome 3 with an average density of one polymorphic site per 2.3 kb (~ 0.14 centimorgans, cM) and establishes a chromosome-wide haplotype for each clone. The combination of MS and SNP markers provides even coverage of the chromosome, which will become a valuable tool for mapping genes associated with various parasite traits.

The predominance of nsSNPs found in *P. falciparum* genes indicates that there may be specific mechanisms affecting the proportion of non-synonymous nucleotide substitutions in this organism. If we calculate the frequency of sSNPs and nsSNPs adjusted for the frequency of synonymous and non-synonymous sites, respectively, we find that the ratio of sSNPs per synonymous

site to nsSNPs per non-synonymous site is $\sim 3:1$ in human and $\sim 2:1$ in *P. falciparum*¹⁴ (Table 1). Furthermore, *P. falciparum* has a lower synonymous substitution rate per synonymous site when compared with human ($\theta_{falciparum}/\theta_{human} = 0.72$), but a higher non-synonymous substitution rate per non-synonymous site ($\theta_{falciparum}/\theta_{human} = 1.17$)¹⁴ (Table 1), although neither comparison is statistically significant. Neutrality tests¹⁷ indicate that the non-synonymous rate is significantly higher than the synonymous rate in only one of the 204 genes (*csp* is the only gene to show evidence of significant positive selection, as reported previously¹⁸). Although the test may have low power because only five sequences were analysed, there does not appear to be exceptionally strong selection acting to increase amino-acid diversity. Thus, it is highly unlikely that the observed chromosome-wide bias against synonymous substitutions is solely due to drug or immune pressures.

The *P. falciparum* genome has an A + T content of $\sim 80\%$ (ref. 13), leading to compositional constraints¹⁹. We found that the genes on chromosome 3 have an average of 19.1% ($\pm 1.5\%$) synonymous sites, of which 7.1% ($\pm 2.2\%$) are four-fold degenerate sites, compared with $\sim 29\%$ and $\sim 16\%$, respectively, in human genes^{14,15}. Lower G + C content at the second codon position but

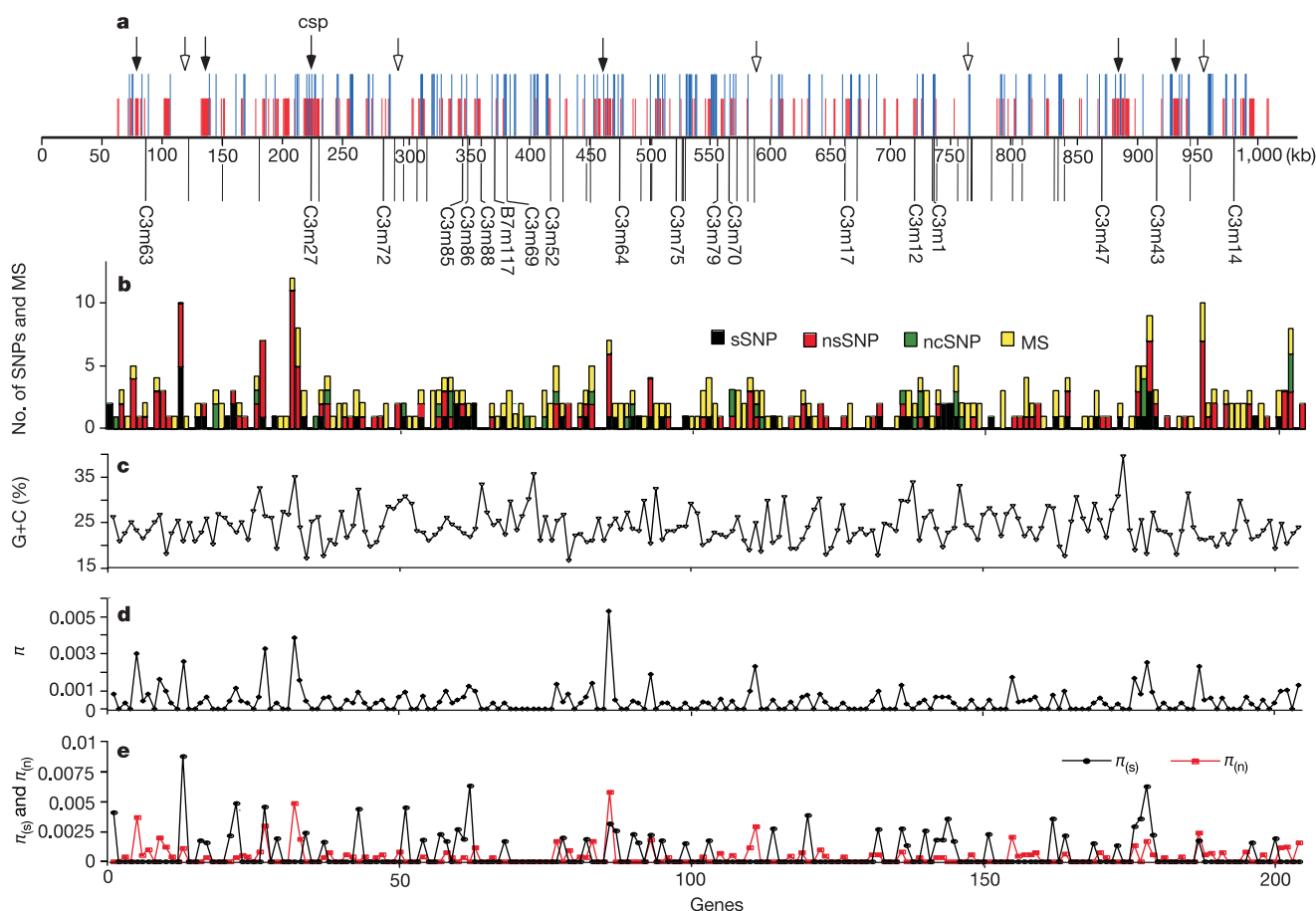


Figure 1 Frequency and distribution of SNPs and MS in 204 genes on chromosome 3 of *P. falciparum*. **a**, A physical map of chromosome 3 integrated with SNPs and MS markers. On top of the chromosome line are SNPs (red) and MS (blue) identified by this work; at the bottom are MS markers and linkage groups reported previously¹⁶. Each named MS marker represents a linkage group established on the basis of genotypes of 35 progeny from a genetic cross¹⁶. SNPs are distributed across the whole chromosome, although there are regions with more SNPs (filled arrows) and regions of up to 4–5 genes without any SNPs (open arrows). A recombination hotspot with five recombination events in ~ 40 kb (between 340 and 380 kb) and two regions of ~ 150 kb with no recombination (C3m63/C3m27 and C3m1/C3m47), are similar to 'jungles' and 'deserts' seen in the

human genome²⁷. **b**, SNPs and MS from individual genes are plotted according to the position of the gene on the chromosome. Most of the genes have 1 or 2 SNPs. Black bars represent synonymous SNPs (sSNPs); red bars, non-synonymous SNPs (nsSNPs); green bars, non-coding SNPs (ncSNPs); yellow bars, MS. **c**, Plots of G + C content from each coding region. **d**, Nucleotide diversity (π) plot from the 204 predicted genes, showing relatively uniform distribution of nucleotide diversity across the chromosome. **e**, Plots of synonymous ($\pi_{(s)}$, black) and non-synonymous ($\pi_{(n)}$, red) nucleotide diversity from each coding region. A higher synonymous substitution rate suggests weak or no functional constraint on synonymous sites.

Table 1 Polymorphisms in 204 genes of chromosome 3

Polymorphism	No. of bp screened	No. of polys	Frequency (SNP/bp)	θ ($\times 10^{-4}$) (mean $\pm$ s.e.)	π ($\times 10^{-4}$) (mean $\pm$ s.e.)
Total SNPs	218,152	238	1/917	5.2	4.9
Non-coding	16,083	31	1/519	10.3 $\pm$ 1.77	9.4 $\pm$ 0.85
Coding	202,069	207	1/976	4.8 $\pm$ 0.59	4.4 $\pm$ 0.57
Synonymous	38,480	62	1/620*	7.2 $\pm$ 1.04*	6.8 $\pm$ 0.97*
Non-synonymous	163,498	145	1/1,128†	4.2 $\pm$ 0.71†	3.9 $\pm$ 0.55†
MS	218,152	165	1/1,299		

Data are obtained from five *P. falciparum* isolates; θ , nucleotide polymorphism; π , nucleotide diversity.

*Calculated based on synonymous substitutions over synonymous sites.

†Adjusted similarly for the frequency of non-synonymous sites.

not the third codon position is also significantly correlated with a lower percentage of synonymous sites (data not shown). If we correct for the reduction in synonymous sites, we would expect to see about 46% sSNPs ($30 \times 29/19$) compared with the observed 30% sSNPs. Clearly, the A + T bias has a role in reducing the numbers of synonymous sites, which in turn leads to the reduced numbers of sSNPs; however, there is no discernable correlation between G + C content and nucleotide mutation rate (Fig. 1c–e).

Another potential explanation for the predominance of nsSNPs is the presence of a high percentage of proteins with low-complexity regions^{12,20} (LCR). LCRs are usually found in protein regions consisting of tandem repeats, representing rapidly evolving non-globular domains, and are thought to be non-essential for protein function and therefore may not be subject to strong selection^{20,21}. To see whether the abundance of LCRs contributes to the bias towards nsSNPs and to test the hypothesis of reduced functional constraints in LCR, we first identified LCRs using the SEG program²¹ (settings: window 45, trigger 3.161 and extension 3.461; J. C. Wootton and O.H.B., manuscript in preparation). We found that 33.8% of amino acids are in LCRs, with 67.2% of the proteins having at least one LCR. Similarly, 32.6% of coding SNPs from chromosome 3 are in LCRs. Although the percentage of nsSNPs is higher in LCRs (76.6%) than that of globular domains (68.3%), the difference is not significant ($P = 0.2358$). If we correct for this bias, we would expect to see 32% sSNPs compared to the observed 30%. These results show that LCRs do not contribute significantly to the bias towards nsSNPs, and suggest that LCR may be under functional constraints similar to the globular regions in this parasite. A combination of drug and immune selection, reduction in synonymous sites due to compositional constraint, and abundance of LCRs may account for the predominance of nsSNPs in the parasite.

To investigate whether there has been a recent severe population bottleneck, we estimated the divergence time of the parasite based on sSNPs and ncSNPs from the whole chromosome. If we used the

synonymous substitution rates estimated assuming that *P. falciparum*/*P. reichenowi* diverged 5–7 million years ago in conjunction with equation (1) in ref. 1, we obtained an estimated divergence time of 126,000–177,000 years ago based on 62 sSNPs (Table 2). Similarly, based on 31 ncSNPs from ~16 kb intron sequences (we used only ncSNPs located in regions of good sequence alignment), we estimated the time of divergence for the parasite to be between 102,000 and 143,000 years ago, using the methods and estimated substitution rates described above^{1,2} (Table 2). Our estimates based on both sSNPs and ncSNPs are similar, and greatly predate the proposed time frame for the Malaria's Eve hypothesis. Indeed, even under the assumption of 7 million years of divergence time between *P. falciparum* and *P. reichenowi*, the lower bounds of the 95% confidence intervals in our estimates (Table 2) are larger than those reported in ref. 1 (0–57,481 years ago) and calculated for data in ref. 2 (4,753–67,237 years ago). Our estimates approximately coincide with the beginning of recent human population expansion¹¹, which could have provided the key element for the growth of the parasite populations.

To estimate the current effective population size N_e , we used the expression $\theta = 2N_e\mu$ and values of μ (1.2×10^{-9} and 1.7×10^{-9}) described in ref. 5. Based on the nucleotide polymorphism observed from non-coding regions ($\theta = 0.00103$), we estimate N_e to be $(3.03\text{--}4.29) \times 10^5$. If we used synonymous nucleotide polymorphism ($\theta = 0.00072$), we obtained $N_e = (2.1\text{--}3.0) \times 10^5$. Our estimates are quite large and of a similar magnitude to those previously reported⁵. These estimates do not support the hypothesis of a severe recent population bottleneck.

There are two plausible explanations for the discrepancy between previous estimates of the age of the parasite and our estimates. First, we genotyped more than 100 worldwide isolates with ~350 MS markers and, based on the genotypes, chose the isolates to maximize diversity²². The five isolates that we used are from different continents and are approximately equally divergent. There was no conspicuous outlier among the five isolates, although the 3D7 is most different from the other four isolates (Fig. 2). On the other hand, the isolates used in previous studies^{1,2} could be closely related⁶. There have been large-scale campaigns using chloroquine and DDT to eradicate malaria in the 1950s, which could have led to drug selective sweeps and reduced diversity in local populations. Second, the sources of DNA sequences and the number of genes

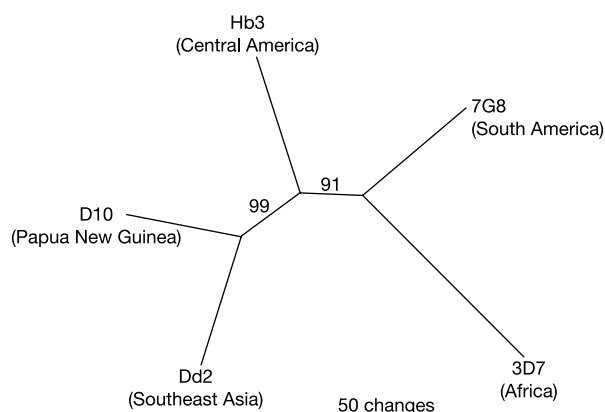


Figure 2 A maximum-likelihood tree based on concatenated DNA sequences of the 204 genes. The tree was constructed using PAUP 4.0 (ref. 28) after obtaining the best-fit maximum-likelihood model²⁹. Bootstrap values were obtained with 1,000 replicates.

Table 2 Estimates of divergence time using ncSNPs and sSNPs from chromosome 3

	μ_a ($\times 10^{-9}$)*	μ_b ($\times 10^{-9}$)*	t_i ($\times 10^5$ yr)†	t_s ($\times 10^5$ yr)
<i>P. reichenowi</i> /7M	2.7	1.33	1.43 (0.97–2.03)‡	1.77 (1.36–2.27)‡
<i>P. reichenowi</i> /5M	3.78	1.86	1.02 (0.69–1.45)‡	1.26 (0.97–1.62)‡

* μ_a and μ_b are the substitution rates at the four- and two-fold degenerate sites derived from estimated divergence time of 5–7 million years between *P. falciparum* and *P. reichenowi*. t_i is the divergence time calculated based on ncSNPs from ~16 kb non-coding sequences; ncSNPs are treated as four-fold degenerate sites as described². t_s is calculated based on sSNPs, including two- and four-fold degenerate sites, from ~202 kb coding sequences.

†Similar results of 130,000–185,000 years were obtained from 34 ncSNPs out of 13.8 kb non-coding sequences of 49 transporter genes located on different chromosomes (data not shown).

‡Numbers in parentheses indicate the 95% confidence interval.

used are important. Previous estimates of parasite diversity were restricted by DNA sequence availability and used a limited number of genes and/or sequences from GenBank. Databases frequently contain erroneous DNA sequences²³, and it is difficult to verify the accuracy of the sequences and the identities of the parasites from which they were obtained. Although synonymous sites and non-coding sequences are assumed to be relatively neutral, the degree of deviation from neutrality may not be the same for genes under various selection pressures. Most of the genes used in previous studies^{1,2} are immune/drug targets or housekeeping genes that could be under strong selection.

In contrast, we have used a large number of SNPs and parasites of highly diverse genetic background to estimate the divergence time and effective population size. We have shown that the *P. falciparum* population is quite ancient and diverse, a result that will be important in the development of drugs and vaccines. □

Methods

Gene sequences and parasite DNA

Predicted coding sequences of the 3D7 parasite were downloaded from PlasmoDB (<http://www.plasmodb.org>). Coding sequences were defined according to annotation in the database. The whole gene sequence, including any introns, was also obtained for sequence alignment. Introns were identified by sequence alignment of predicted coding sequences and the genomic DNA sequence. Parasite culture and DNA extraction were as described²⁴.

Amplification of parasite DNA and DNA sequencing

Two to six primers were designed to amplify and sequence DNA fragments up to 1.5 kb from each of 215 predicted genes on chromosome 3, excluding multigene families such as *rifin*, *var*, *stevo* and *clag*¹⁵. PCR set-ups contained 4 µl DNA (~5 ng), 0.3 µl of each primer (100 µM), and 45 µl of PCR mix containing 5 µl 10X PCR buffer, 1.0 µl dNTPs (10 mM), and 0.2 µl (5 U µl⁻¹) *Taq* polymerase (Roche Applied Science). All amplifications were performed with one cycling condition: 94 °C for 2 min, 35 cycles of 94 °C for 20 s, 52 °C for 10 s, 48 °C for 10 s, and 60 °C for 1.5 min. The PCR product was treated with 2 µl ExoSAP-IT (USB) at 37 °C for 15 min and 80 °C for another 15 min. 5 µl of the treated product was used in sequencing reaction with dichlororhodamine or BigDye terminator chemistry on an ABI377 or ABI3100 DNA sequencer (Applied Biosystem).

Data analysis

DNA sequences were aligned using Sequencher 3.1 (Gene Codes Corp.). All potential SNPs and discrepancies were verified by visual inspection. The DnaSP program²⁵ was used to calculate π , θ , and other parameters for each gene. The neutrality test was done using the codonml program of the software package PAML¹⁷. Estimates of nucleotide substitution rates, divergence time, and population size were made using methods described elsewhere^{1,2,5}. To obtain LCR from the genes, we first corrected for the background amino acid composition bias by randomly shuffling all the residues and determined the trigger complexity that results in 4% of the database being within LCR²⁶ (J. C. Wootton and O.H.B., manuscript in preparation). We then used this trigger complexity to run the SEG program²¹.

Received 21 January; accepted 24 April 2002; doi:10.1038/nature00836.

- Rich, S. M., Licht, M. C., Hudson, R. R. & Ayala, F. J. Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **95**, 4425–4430 (1998).
- Volkman, S. K. *et al.* Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* **293**, 482–484 (2001).
- Hughes, A. L. & Verra, F. Ancient polymorphism and the hypothesis of a recent bottleneck in the malaria parasite *Plasmodium falciparum*. *Genetics* **150**, 511–513 (1998).
- Hey, J. Parasite populations: the puzzle of *Plasmodium*. *Curr. Biol.* **9**, R565–R567 (1999).
- Hughes, A. L. & Verra, F. Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proc. R. Soc. Lond. B* **268**, 1855–1860 (2001).
- Saul, A. Circumsporozoite polymorphisms, silent mutations and the evolution of *Plasmodium falciparum*. *Parasitol. Today* **15**, 38–40 (1999).
- Escalante, A. A., Barrio, E. & Ayala, F. J. Evolutionary origin of human and primate malaria: evidence from the circumsporozoite protein gene. *Mol. Biol. Evol.* **12**, 616–626 (1995).
- Walliker, D., Babiker, H. & Ranford-Cartwright, L. in *Malaria: Parasite Biology, Pathogenesis, and Protection* (ed. Sherman, I. W.) 235–252 (American Society for Microbiology, Washington DC, 1998).
- Conway, D. J. *et al.* High recombination rate in natural populations of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **96**, 4506–4511 (1999).
- Anderson, T. J. *et al.* Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17**, 1467–1482 (2000).
- Templeton, A. Out of Africa again and again. *Nature* **416**, 45–51 (2002).
- Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).
- Bowman, S. *et al.* The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**, 532–538 (1999).
- Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* **22**, 231–238 (1999).
- Graur, D. & Li, W.-h. *Fundamentals of Molecular Evolution* (Sinauer, Sunderland, Massachusetts, 2000).

- Su, X. *et al.* A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* **286**, 1351–1353 (1999).
- Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
- Hughes, A. L. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. *Genetics* **127**, 345–353 (1991).
- Musto, H., Romero, H., Zavala, A., Jabbari, K. & Bernardi, G. Synonymous codon choices in the extremely GC-poor genome of *Plasmodium falciparum*: compositional constraints and translational selection. *J. Mol. Evol.* **49**, 27–35 (1999).
- Pizzi, E. & Frontali, C. Low-complexity regions in *Plasmodium falciparum* proteins. *Genome Res.* **11**, 218–229 (2001).
- Wootton, J. C. Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput. Chem.* **18**, 269–285 (1994).
- Wootton, J. C. *et al.* Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature*, **418**, 320–323.
- Louis, A., Ollivier, E., Aude, J. C. & Risler, J. L. Massive sequence comparisons as a help in annotating genomic sequences. *Genome Res.* **11**, 1296–1303 (2001).
- Su, X., Kirkman, L. A., Fujioka, H. & Welles, T. E. Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant *P. falciparum* in Southeast Asia and Africa. *Cell* **91**, 593–603 (1997).
- Rozas, J. & Rozas, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175 (1999).
- Wootton, J. C. & Federhen, S. Statistics of local complexity in amino acid sequence databases. *Comput. Chem.* **17**, 149–163 (1993).
- Yu, A. *et al.* Comparison of human genetic and sequence-based physical maps. *Nature* **409**, 951–953 (2001).
- Swofford, D. L. *PAUP**. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4 (Sinauer Associates, Sunderland, Massachusetts, 2000).
- Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).

Acknowledgements

We thank T. Anderson, A. Saul, K. Hayton and J. Ribeiro for critical reading of the manuscript, and B. Marshall for editorial assistance. We also thank J. Wootton for supporting O.H.B. This work was partially supported by the NIH (W.-H.L.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.-z.S. (e-mail: xsu@niaid.nih.gov).

Dendritic spikes as a mechanism for cooperative long-term potentiation

Nace L. Golding*†, Nathan P. Staff† & Nelson Spruston

Department of Neurobiology and Physiology, Institute for Neuroscience, Northwestern University, 2153 North Campus Drive, Evanston, Illinois 60208-3520, USA

† These authors contributed equally to this work

Strengthening of synaptic connections following coincident pre- and postsynaptic activity was proposed by Hebb as a cellular mechanism for learning¹. Contemporary models assume that multiple synapses must act cooperatively to induce the postsynaptic activity required for hebbian synaptic plasticity^{2–5}. One mechanism for the implementation of this cooperation is action potential firing, which begins in the axon, but which can influence synaptic potentiation following active backpropagation into dendrites⁶. Backpropagation is limited, however, and action potentials often fail to invade the most distal dendrites^{7–10}. Here we show that long-term potentiation of synapses on the distal dendrites of hippocampal CA1 pyramidal neurons does require cooperative synaptic inputs, but does not require axonal action potential firing and backpropagation. Rather, locally

* Present address: Section of Neurobiology, University of Texas at Austin, Patterson Labs, Austin, Texas 78712, USA

generated and spatially restricted regenerative potentials (dendritic spikes) contribute to the postsynaptic depolarization and calcium entry necessary to trigger potentiation of distal synapses. We find that this mechanism can also function at proximal synapses, suggesting that dendritic spikes participate generally in a form of synaptic potentiation that does not require postsynaptic action potential firing in the axon.

Hebb's view of the malleable synapse¹ has gained tremendous experimental support. Repeated pairing of presynaptic stimulation with postsynaptic depolarization leads reliably to long-term potentiation (LTP) of excitatory postsynaptic potentials (EPSPs)^{11–13}. Postsynaptic action potentials can provide the necessary depolarization at some dendritic synapses, via backpropagation from the axon into dendrites⁶. At the most distal synapses, however, this mechanism may be ineffective, because action potential amplitude decreases markedly between the axon and the distal dendrites, especially during the repetitive firing necessary to induce LTP^{7–10}. An alternative postsynaptic signal could be provided by dendritically generated spikes, which are produced by activation of sodium and calcium channels by large synaptic potentials in hippocampal or cortical dendrites^{14–20}.

To test this idea, we activated synapses on the distal dendrites of CA1 pyramidal neurons by placing a stimulating electrode in stratum lacunosum-moleculare (SLM) of hippocampal slices prepared from mature rats (Methods). LTP was induced using a physiologically relevant stimulus^{6,21} consisting of EPSPs (2–4 mV) stimulated in bursts of five and paired with three postsynaptic action potentials, repeated at the naturally occurring theta frequency (Methods; Fig. 1a and b). To investigate the role of backpropagating action potentials in this form of LTP, tetrodotoxin (TTX) was applied to the axon, soma, and proximal dendrites. This

manipulation prevented action potential initiation and backpropagation during the theta-burst stimulus, without affecting evoked EPSPs.

LTP at distal synapses was not affected by blocking action potential firing with proximal application of TTX (proximal TTX; Fig. 1c). In addition, LTP was not prevented by somatic hyperpolarization or voltage clamp near the resting potential, manipulations that prevented action potential initiation but which should not prevent depolarization of the distal dendrites²² (Fig. 1e). Although backpropagating action potentials were not required, LTP at distal synapses did require significant depolarization, as smaller EPSPs (0.5–1 mV, paired with action potential firing) did not produce LTP, indicating that multiple synapses must cooperate to produce sufficient postsynaptic depolarization^{2–5}. LTP could be observed at proximal synapses (with the stimulating electrode placed in stratum radiatum, SR) when small EPSPs were paired with action potentials using the theta-burst protocol (Fig. 1d and e). Thus, synaptic inputs sufficient to trigger axonally initiated, backpropagating action potentials can account for the cooperativity of LTP at proximal, but not distal, synapses.

We proposed that EPSPs summing during theta-burst stimulation of SLM produced cooperative LTP owing to the occurrence of regenerative spikes initiated in the dendrites. To test this hypothesis, we limited dendritic spiking by blocking a subset of voltage-activated calcium channels, while preventing backpropagating action potentials with proximal application of TTX. A combination of 50 μ M nickel and 10 μ M nimodipine was used to block a fraction of the low-threshold and high-threshold channels (primarily T- and L-type channels, respectively). Blocking some calcium channels in this way reduced the average magnitude of LTP from an 85% increase to a 52% increase (Fig. 2a and b), suggesting that activation

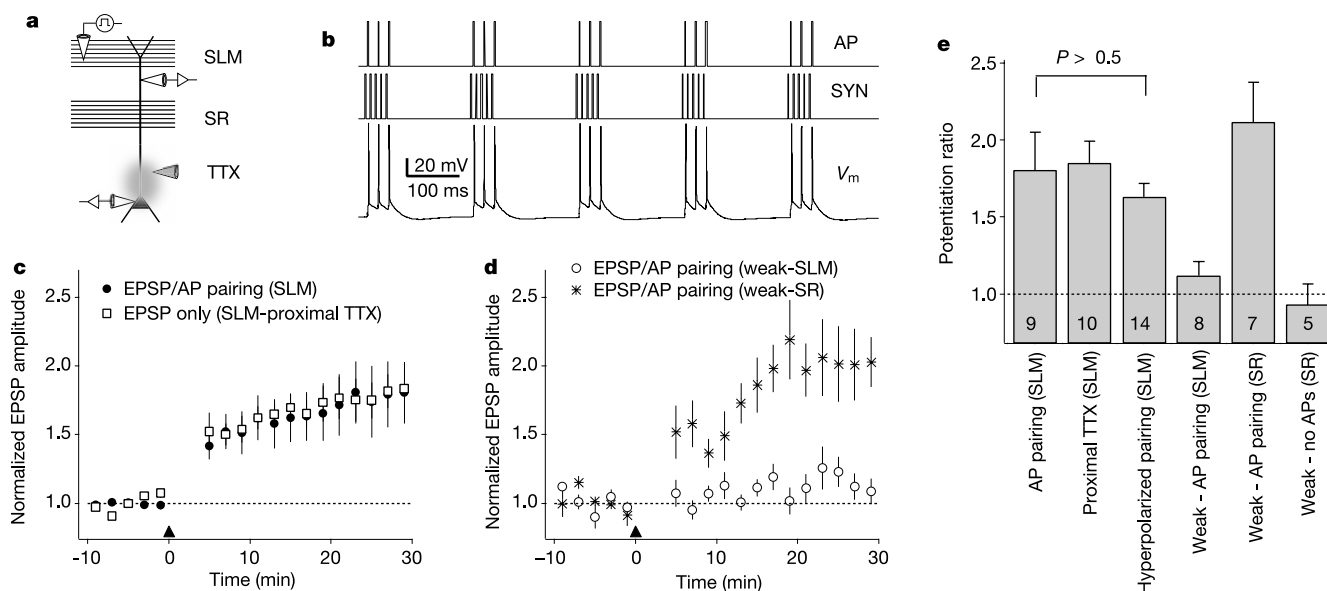


Figure 1 Induction of long-term potentiation at distal synapses is independent of backpropagating action potentials. **a**, Recording configuration with whole-cell somatic patch-clamp recording, stimulation of stratum lacunosum-moleculare (SLM) axons, and application of 5 μ M tetrodotoxin (TTX) near the soma to block action potentials in some experiments. SR, stratum radiatum. **b**, Theta-burst LTP induction protocol. Bursts of EPSPs (SYN; 5 $\times$ synaptic stimulation at 100 Hz) were paired with triplets of action potentials (AP; 2-ms somatic current steps at 50 Hz). A single theta stimulus consisted of five EPSP/AP bursts at 5 Hz (V_m ; one postsynaptic voltage response is shown). The theta stimulus was repeated 3 times, separated by 30 s. **c**, Normalized EPSP amplitude (2–4 mV) before and after theta-burst stimulation shows a similar degree of potentiation

when induced with (filled circles) and without (proximal TTX, open boxes) the pairing of action potentials. **d**, Normalized EPSP amplitude (0.5–1 mV) before and after theta-burst stimulation. Weak stimulation paired with action potentials does not induce LTP at SLM synapses (open circles), but does at SR synapses (asterisks). **e**, Potentiation ratio (average EPSP 20–30 min after theta-burst stimulation to the average of the 10-min baseline) under various experimental conditions. Hyperpolarized pairing includes bursts paired with a 100-ms hyperpolarizing pulse (–1.0 to –1.5 nA) or somatic voltage clamp to the resting potential (about –65 mV). The number of experiments in each condition is indicated at bottom of each bar.

of voltage-gated calcium channels contributes to LTP at SLM synapses. A similar reduction in LTP was observed (from 85% increase to a 57% increase) when NMDA- (*N*-methyl-D-aspartate) type glutamate receptors were blocked by application of 50 μ M AP5 and 20 μ M MK-801 (Fig. 2a and b). Robust block of LTP was obtained when both NMDA receptors and a fraction of voltage-gated calcium channels were blocked together (Fig. 2a and b). In each of these experiments, the drugs were continuously present throughout the experiment, and baseline EPSP amplitude was set in the 2–4 mV range. There were no differences in the average baseline EPSP amplitude across the four groups. Separate experiments also indicated that these drugs did not affect peak amplitude or initial slope of individual EPSPs ($n = 3$, see Supplementary Information). The observed effects on LTP, however, suggest that bursts of EPSPs in SLM activate both NMDA receptors and calcium channels, which act synergistically to produce the local depolarization and calcium influx necessary to induce LTP.

To test further the hypothesis that the depolarization achieved during theta-burst stimulation was a critical determinant of co-operative LTP at SLM synapses, we examined the relationship between the magnitude of LTP and the change in somatic membrane potential (V_m) during the theta-burst stimulus (average peak ΔV_m during the 15 bursts used to induce LTP). Pooling the data for all four conditions (proximal TTX alone, or with partial calcium channel block, NMDA receptor block, or partial calcium channel and NMDA receptor block), a strong correlation emerged between the amount of depolarization and the magnitude of LTP (Fig. 2c). By contrast, no correlation was observed between the size of the baseline EPSP (2–4 mV) and the amount of LTP (Fig. 2d). These data suggest that EPSP summation and active responses during theta bursts are more important determinants of LTP than initial

synaptic strength.

Sample traces during theta-burst stimuli are shown in Fig. 2e (examples from three experiments for each condition). Many of the responses exhibited signs of regenerative activity in the somatic recordings, despite the presence of somatically applied TTX. The only exception was when calcium channels and NMDA receptors were blocked together, in which case signs of regenerative activity were almost never observed (and there was less EPSP summation and little LTP). As the presence of regenerative events markedly increased the peak somatic depolarization during theta-burst stimulation, we proposed that these events contribute to the induction of LTP. Further, we postulated that these events were generated in the distal dendrites, where the activated synapses were located. To examine the correlation between these putative dendritic spikes and the induction of LTP, each experiment was analysed for the presence of regenerative events (rapid depolarizations greater than 3 mV above a sharp inflection point). Comparing experiments with and without putative dendritic spikes (filled and open symbols or bars in Figs 2c, d and f), much more LTP was observed when regenerative spikes were recorded in the soma than when they were absent.

To test directly the hypothesis that the somatic regenerative events reflected dendritically generated spikes, we performed simultaneous recordings from the soma and distal dendrites of CA1 neurons ($n = 6$; beyond 300 μ m from the soma). Regenerative events were recorded in the dendrites during theta-burst stimulation of SLM (Fig. 3A). The largest of these events triggered somatic action potentials under normal conditions, but smaller events failed to trigger action potentials (Fig. 3A, a). When TTX was applied locally to the soma and proximal dendrites, dendritic spikes were always much larger than the correlated events in the soma (Fig. 3A,

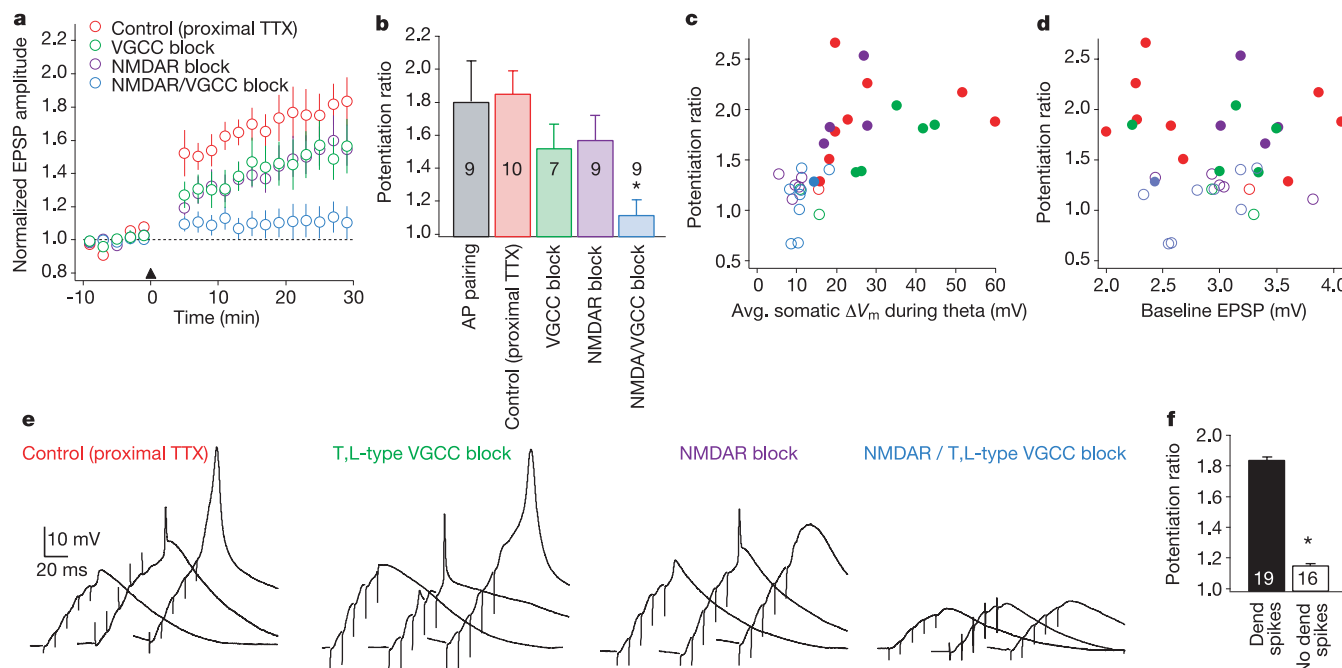


Figure 2 LTP at distal synapses requires voltage-gated calcium channels and NMDA receptors. **a**, Normalized EPSP amplitude in response to SLM stimulation before and after theta-burst stimulus under different pharmacological conditions. Action potentials were always suppressed with proximal TTX application. Conditions: control (proximal TTX alone), NMDA receptor (NMDAR) blockade with 50 μ M AP-5 and 20 μ M MK-801, blockade of T- and L-type calcium channels (VGCC) with 50 μ M NiCl₂ and 10 μ M nimodipine, and combined block of both NMDA receptors and T- and L-type calcium channels. **b**, Potentiation ratios induced by theta-burst stimuli applied to SLM in the different pharmacological conditions (asterisk, $P < 0.05$; the number of experiments for

each condition is indicated for each bar). **c**, Magnitude of LTP is correlated with the amplitude of the theta-burst stimulus response at the soma ($r = 0.62$; $P < 0.0001$). Symbols indicate cells where dendritic spikes were (filled) or were not (open) observed at the soma. **d**, Magnitude of LTP is not correlated with the amplitude of the baseline EPSP over a range spanning 2–4 mV ($r = -0.03$). Symbols as in c. **e**, Examples of theta-burst stimulus responses in the different pharmacological conditions. Spikes are apparent in all conditions except the combined NMDA receptor and calcium channel blockers. **f**, Magnitude of LTP is significantly greater in cells exhibiting dendritic spikes recorded at the soma as compared to those without visible dendritic spikes ($P < 10^{-6}$).

b)^{15,17,23}. These double recordings show that the small regenerative events recorded in the soma correspond to larger, dendritically generated events that propagate poorly toward the soma. The correlation between these regenerative events and LTP strongly suggests that dendritic spikes contribute to LTP induction (Fig. 2c, d and f). Furthermore, dendritic spikes were sometimes present even when the somatic recording showed no obvious signs of regenerative activity. Such events may contribute to the LTP that was sometimes observed when no dendritic spikes were apparent in the somatic recording (Fig. 2c, d and f).

The contribution of dendritic spikes to distal dendritic depolarization and calcium entry was studied further using calcium imaging experiments. We compared the amplitude of distal dendritic calcium transients in SLM produced by a single burst of EPSPs just below and just above threshold for dendritic spikes detected at the soma (Fig. 3B, a–c). Despite the fact that these EPSPs were almost identical before the dendritic spike (Fig. 3B, b), the average calcium transient elicited with dendritic spikes apparent at the soma

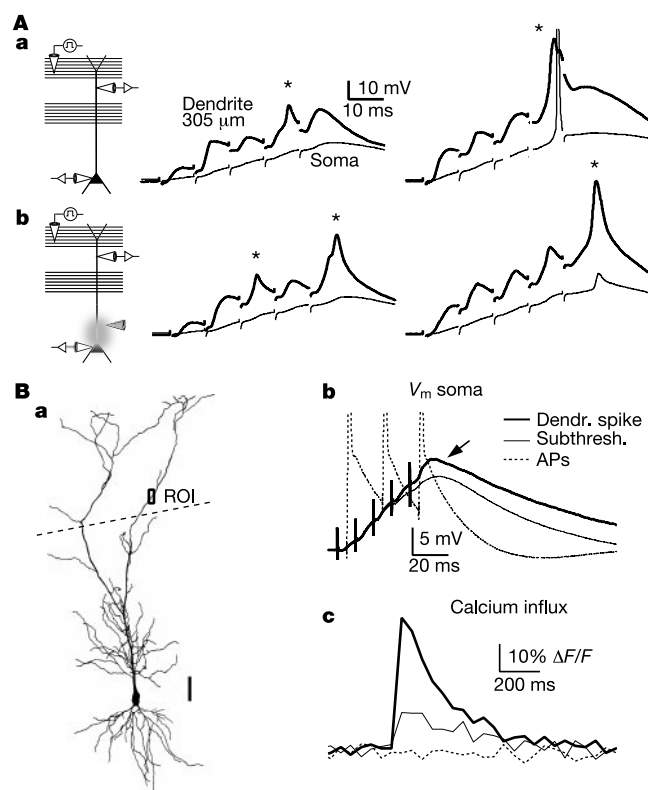


Figure 3 Large theta-burst responses and distal calcium influx are associated with dendritic spikes. **A,a**, Theta-burst responses recorded simultaneously at the soma and distal dendrite. Dendritic spikes (asterisks) appeared highly attenuated at the soma, and were variably coupled to action potential initiation in the axon. **A,b**, Proximal application of TTX blocked somatic action potentials without affecting dendritic spikes (asterisks) or synaptic potentials. Small spikes at the soma correspond to much larger spikes in the dendrites. **B,a**, Reconstruction of biocytin-labelled neuron used for calcium imaging. Calcium transients were imaged in the SLM, 430 μm from the soma. For clarity, the box indicating the region of interest (ROI) is exaggerated in size. Dotted line indicates the SR/SLM border. Scale bar, 50 μm. **B,b**, The cell was stimulated either with 50 Hz triplets of action potentials (dotted trace) or with single burst synaptic stimuli evoked at intensities just below (thin trace) and above threshold (thick trace) for dendritic spike initiation (detected in the somatic recording as a small spikelet, arrow). **B,c**, Calcium influx (reported by the relative fluorescence change, $\Delta F/F$) in SLM during the three stimuli shown in **B,b**. Dendritic spikes were associated with large calcium transients ($\Delta F/F$, see Methods), whereas backpropagating action potentials and subthreshold theta-burst responses were associated with smaller calcium accumulations.

was 2.63 ± 0.45 times that without dendritic spikes ($n = 6$ cells; range 1.47–4.25; Fig. 3B, c). In contrast to the large calcium transients evoked by dendritic spikes, triplets of backpropagating action potentials did not evoke calcium transients above the noise in the distal dendrites of any of the five cells tested (for example, Fig. 3B, c). These findings suggest that large, dendritically initiated spikes produce much greater depolarization and calcium entry in distal CA1 dendrites than synaptic potentials (either alone or possibly with smaller dendritic spikes not detected at the soma) or backpropagating action potentials.

Taken together, our results strongly implicate dendritically generated spikes as key events in the induction of LTP at distal synapses on CA1 neurons, where action potential backpropagation is limited and does not affect LTP. We asked further whether dendritic spikes could also induce LTP at more proximal synapses on CA1 neurons. To test this idea, we induced LTP by stimulating SR strongly during the theta bursts while applying TTX onto the axon, soma, and proximal dendrite to prevent action potential initiation during theta-burst stimulation. Robust LTP was observed under these conditions, and regenerative activity clearly had its origin in the dendrite ($n = 4$, Fig. 4).

Voltage-gated calcium channels and NMDA receptors act synergistically to induce LTP during theta-burst stimulation (Fig. 2; see also ref. 6). This can be attributed to the ability of each of these channels to produce depolarization and voltage-dependent calcium influx²⁴. The occurrence of LTP in the presence of NMDA receptor antagonists suggests that NMDA receptor activation facilitates this form of LTP, but is not required. The persistence of dendritic spikes and LTP during partial block of voltage-gated calcium channels suggests that the sodium channels and unblocked calcium channels can act together with AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) and NMDA receptor activation to induce LTP. NMDA receptor activation may also contribute to the dendritic spikes that enhance LTP²⁵.

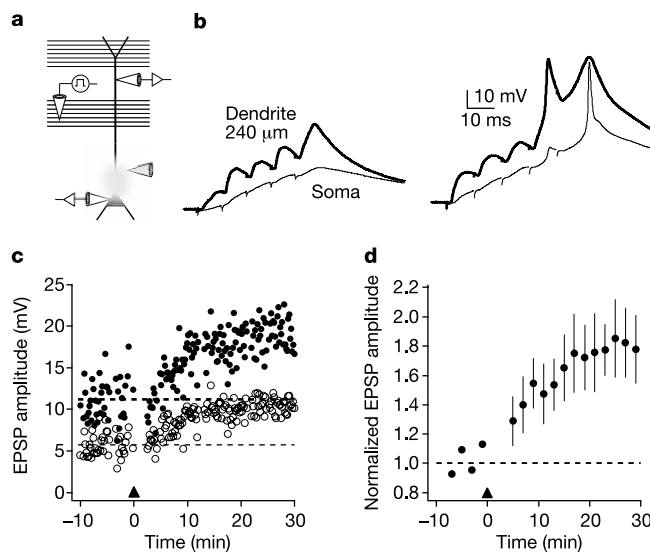


Figure 4 Dendritic spikes contribute to LTP induction at stratum radiatum (SR) synapses. **a**, Experimental configuration. Responses to SR stimulation were recorded in the dendrites and soma. TTX was applied near the soma during theta-burst stimulation. **b**, Examples of dendritic spikes in response to theta-burst stimulation in SR. Dendritic spikes are strongly attenuated at the soma. **c**, LTP induction at SR synapses can occur independently of backpropagating action potentials. EPSPs measured simultaneously in the dendrite (filled circles) and soma (open circles) before and after theta-burst stimulation in the presence of somatic TTX. **d**, Summary data of dendritic EPSP amplitudes in LTP experiments where backpropagating action potentials were blocked ($n = 4$).

Our findings reveal an important function of dendritically generated spikes. Although these events may contribute to action potential initiation in the axon, their propagation from the dendrites to the soma is limited, both *in vitro* and *in vivo*^{9,14–18}, suggesting that they may also serve a local function in dendrites. Our data suggest that dendritic spikes can provide the depolarization and calcium entry required for LTP. Although backpropagating action potentials had no effect on theta-burst-induced LTP at SLM synapses, they may be important at these synapses under other conditions. For example, backpropagating action potentials may alter the threshold for dendritic spikes and the extent of their propagation within the dendritic tree²⁶.

Although we used somatic TTX application or hyperpolarization to prevent action potential initiation and backpropagation during theta-burst stimulation, a similar situation may occur *in vivo* if the soma is inhibited while the dendrites are excited. Indeed, during hippocampal theta rhythm the distal dendrites are excited by glutamate-mediated synaptic potentials, which can trigger dendritic spikes, at a time when the soma is inhibited by GABA- (γ -aminobutyric acid) mediated synaptic potentials^{16,27}. Dendritic inhibition may also influence dendritic spikes and LTP. Inhibition that arrives concurrently with excitation would limit dendritic spikes²⁸, whereas inhibition that arrives before an excitatory input could enhance dendritic spikes by removing inactivation of sodium and calcium channels.

The existence of a form of LTP that does not require action potential output via the axon has implications for plasticity rules and the cellular mechanisms of learning. Whereas synapses located anywhere in the dendrites can cooperate to produce action potentials in the axon and backpropagation-induced LTP, only spatially co-localized synapses are expected to cooperate to produce dendritic-spike-induced LTP. Also, although backpropagating action potentials invade much of the dendritic tree¹⁰, the spread of dendritic spikes is more limited^{15,17,19}, implying that dendritic spikes will induce LTP at synapses within smaller dendritic domains. The implementation of cooperative LTP^{2–5} via dendritic spikes may therefore lead to the selection of clusters of effective synapses in dendritic domains or branches, a property of dendritic integration that has been proposed to increase the computational power and memory capacity of some neural systems^{29,30}. □

Methods

Hippocampal slice preparation

Transverse hippocampal slices (300 μ m) were made from 5–9-week-old Wistar rats as described previously¹⁵. Dissection and slicing were performed in ice cold, oxygenated (bubbled with 95% O₂, 5% CO₂) artificial cerebrospinal fluid (ACSF), containing (in mM): 125 NaCl, 2.5 KCl, 2 CaCl₂, 1 MgCl₂, 25 NaHCO₃, 1.25 NaH₂PO₄, 25 glucose. Slices were incubated for 30 min in a holding chamber at 35 °C and then stored at room temperature for at least 30 min. Individual slices were then transferred to a submerged recording chamber perfused with ACSF at 35 $\pm$ 2 °C. Neurons were visualized with differential interference contrast optics. Synaptic stimulation was achieved with a relatively large (~20 μ m diameter) patch pipette connected to a stimulus isolator (Axon Instruments), and placed at least 1 mm from the recorded cell either in SLM or the distal third of SR. All experiments were performed in the presence of the GABA_A and GABA_B receptor antagonists SR95531 (4 μ M) and CGP55845A (1 μ M), which facilitated activation of dendritic spikes and induction of LTP^{4,15,17}. Dendritic spikes could also be elicited with half-maximal block of GABA-mediated transmission (0.44 μ M SR95531 and 6 nM CGP55845A), but plasticity was difficult to study owing to contamination of the postsynaptic potential by direct activation of GABA-containing axons (see Supplementary Information).

Electrophysiology

Whole-cell current-clamp recordings were made from cells in conjunction with bridge amplifiers (Dagan BVC-700). Patch-clamp electrodes were fire-polished and filled with intracellular solution containing (in mM): potassium gluconate 115, KCl 20, sodium phosphocreatine 10, HEPES 10, MgATP 2, NaGTP 0.3, and 0.1% biocytin for subsequent determination of morphology. Electrode resistance in the bath ranged from 2–4 M Ω for somatic and 7–10 M Ω for dendritic pipettes, and series resistance ranged from 9 to 50 M Ω . Electrophysiological traces were digitized via an ITC 16 or ITC18 digital-analog converter (Instrutech) under control of macros custom programmed in IGOR Pro (Wavemetrics).

For LTP experiments, EPSP amplitude was monitored using 0.1-Hz synaptic stimulation. LTP was induced using a theta-burst pairing protocol (Fig. 1). When

noted, 5 μ M TTX + 0.1% Fast Green dissolved in ACSF was pressure-applied through a patch pipette to the proximal apical dendrite and soma under visual guidance to prevent axonal action potential generation. All other drugs were bath-applied, and present throughout the entire experiment.

Calcium imaging

In calcium imaging experiments, cells were filled with 150 μ M calcium green-1 (Molecular Probes) via the recording electrode for at least 30 min, and excited at 480 nm wavelength. Fluorescence was averaged from small dendritic regions of interest in the SLM (~3 $\times$ 10 μ m, viewed at 160 $\times$), filtered at 535 nm wavelength, and collected with a frame-transfer, cooled CCD camera (EBFT512; Roper Scientific) under the control of custom macros programmed using IGOR Pro. Time-dependent fluorescence changes were background subtracted, and then calculated as $\Delta F/F$, where ΔF is the time-dependent changes in fluorescence and F is the resting fluorescence. All responses were averages of several trials with a single burst of EPSPs repeated at low frequency (<0.1 Hz). Multiple bursts at theta frequency always evoked larger responses, indicating that the dye was never saturated by a single-burst response.

Data analysis

Analysis of electrophysiology and imaging data was performed using IGOR Pro Software and statistical tests were performed using Excel software (Microsoft). Statistical analysis of multi-group data was performed using a single-factor analysis of variance. When there was a significant difference between the groups, Tukey's multiple comparison tests were performed to determine the level of significance for each pairwise comparison. All measurements are presented as mean $\pm$ s.e.m.

Received 4 April; accepted 9 May 2002; doi:10.1038/nature00854.

- Hebb, D. O. *Organization of Behavior* (Wiley, New York, 1949).
- McNaughton, B. L., Douglas, R. M. & Goddard, G. V. Synaptic enhancement in fascia dentata: cooperativity among coactive afferents. *Brain Res.* **157**, 277–293 (1978).
- Lee, K. S. Cooperativity among afferents for the induction of long-term potentiation in the CA1 region of the hippocampus. *J. Neurosci.* **3**, 1369–1372 (1983).
- Colbert, C. M. & Levy, W. B. Long-term potentiation of perforant path synapses in hippocampal CA1 in vitro. *Brain Res.* **606**, 87–91 (1993).
- Sjostrom, P. J., Turrigiano, G. G. & Nelson, S. B. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* **32**, 1149–1164 (2001).
- Magee, J. C. & Johnston, D. A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science* **275**, 209–213 (1997).
- Spruston, N., Schiller, Y., Stuart, G. & Sakmann, B. Activity-dependent action potential invasion and calcium influx into hippocampal CA1 dendrites. *Science* **268**, 297–300 (1995).
- Callaway, J. C. & Ross, W. N. Frequency-dependent propagation of sodium action potentials in dendrites of hippocampal CA1 pyramidal neurons. *J. Neurophysiol.* **74**, 1395–1403 (1995).
- Stuart, G., Schiller, J. & Sakmann, B. Action potential initiation and propagation in rat neocortical pyramidal neurons. *J. Physiol.* **505**, 617–632 (1997).
- Golding, N. L., Kath, W. L. & Spruston, N. Dichotomy of action-potential backpropagation in CA1 pyramidal neuron dendrites. *J. Neurophysiol.* **86**, 2998–3010 (2001).
- Gustafsson, B., Wigström, H., Abraham, W. C. & Huang, Y.-Y. Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials. *J. Neurosci.* **7**, 774–780 (1987).
- Markram, H., Lubke, J., Frotscher, M. & Sakmann, B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**, 213–215 (1997).
- Bi, G. Q. & Poo, M. M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18**, 10464–10472 (1998).
- Schiller, J., Schiller, Y., Stuart, G. & Sakmann, B. Calcium action potentials restricted to distal apical dendrites of rat neocortical pyramidal neurons. *J. Physiol.* **505**, 605–616 (1997).
- Golding, N. L. & Spruston, N. Dendritic sodium spikes are variable triggers of axonal action potentials in hippocampal CA1 pyramidal neurons. *Neuron* **21**, 1189–1200 (1998).
- Kamondi, A., Acsády, L. & Buzsáki, G. Dendritic spikes are enhanced by cooperative network activity in the intact hippocampus. *J. Neurosci.* **18**, 3919–3928 (1998).
- Golding, N. L., Jung, H. Y., Mickus, T. & Spruston, N. Dendritic calcium spike initiation and repolarization are controlled by distinct potassium channel subtypes in CA1 pyramidal neurons. *J. Neurosci.* **19**, 8789–8798 (1999).
- Helmchen, F., Svoboda, K., Denk, W. & Tank, D. W. In vivo dendritic calcium dynamics in deep-layer cortical pyramidal neurons. *Nature Neurosci.* **2**, 989–996 (1999).
- Häusser, M., Spruston, N. & Stuart, G. J. Diversity and dynamics of dendritic signalling. *Science* **290**, 739–744 (2000).
- Wei, D. S. *et al.* Compartmentalized and binary behaviour of terminal dendrites in hippocampal pyramidal neurons. *Science* **293**, 2272–2275 (2001).
- Pike, F. G., Meredith, R. M., Olding, A. W. & Paulsen, O. Postsynaptic bursting is essential for 'Hebbian' induction of associative long-term potentiation at excitatory synapses in rat hippocampus. *J. Physiol.* **518**, 571–576 (1999).
- Spruston, N., Jaffe, D. B. & Johnston, D. Dendritic attenuation of synaptic potentials and currents: the role of passive membrane properties. *Trends Neurosci.* **17**, 161–166 (1994).
- Segue, I. & Rall, W. Excitable dendrites and spines: earlier theoretical insights elucidate recent direct observations. *Trends Neurosci.* **21**, 453–460 (1998).
- Schiller, J., Schiller, Y. & Clapham, D. E. NMDA receptors amplify calcium influx into dendritic spines during associative pre- and postsynaptic activation. *Nature Neurosci.* **1**, 114–118 (1998).
- Schiller, J., Major, G., Koester, H. J. & Schiller, Y. NMDA spikes in basal dendrites of cortical pyramidal neurons. *Nature* **404**, 285–289 (2000).
- Larkum, M. E., Zhu, J. J. & Sakmann, B. A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature* **398**, 338–341 (1999).
- Kamondi, A., Acsády, L., Wang, X. J. & Buzsáki, G. Theta oscillations in somata and dendrites of hippocampal pyramidal cells in vivo: activity-dependent phase-precession of action potentials. *Hippocampus* **8**, 244–261 (1998).

28. Miles, R., Toth, K., Gulyas, A. I., Hajos, N. & Freund, T. F. Differences between somatic and dendritic inhibition in the hippocampus. *Neuron* **16**, 815–823 (1996).
29. Mel, B. W. Synaptic integration in an excitable dendritic tree. *J. Neurophysiol.* **70**, 1086–1101 (1993).
30. Poirazi, P. & Mel, B. W. Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron* **29**, 779–796 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

This work was supported by the National Institutes of Health.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.S. (e-mail: spruston@northwestern.edu).

Identification of genes expressed in *C. elegans* touch receptor neurons

Yun Zhang*, Charles Ma*, Thomas Delohery†‡, Brian Nasipak*, Barrett C. Foat*, Alexander Bounoutas*, Harmen J. Bussemaker*, Stuart K. Kim§ & Martin Chalfie*

* Department of Biological Sciences, Columbia University, New York, New York 10027, USA

† Memorial Sloan Kettering Institute, New York, New York 10021, USA

§ Department of Developmental Biology and Genetics, Stanford University Medical Center, Stanford, California 94305, USA

The extent of gene regulation in cell differentiation is poorly understood. We previously used saturation mutagenesis to identify 18 genes that are needed for the development and function of a single type of sensory neuron—the touch receptor neuron for gentle touch in *Caenorhabditis elegans*^{1,2}. One of these genes, *mec-3*, encodes a transcription factor that controls touch receptor differentiation^{3,4}. By culturing and isolating wild-type and *mec-3* mutant cells from embryos and applying their amplified RNA to DNA microarrays, here we have identified genes that are known to be expressed in touch receptors, a previously uncloned gene (*mec-17*) that is needed for maintaining touch receptor differentiation^{2,5}, and more than 50 previously unknown *mec-3*-dependent genes. These genes are randomly distributed in the genome and under-represented both for genes that are co-expressed in operons and for multiple members of gene families. Using regions 5' of the start codon of the first 20 genes, we have also identified an over-represented heptanucleotide, AATGCAT, that is needed for the expression of touch receptor genes⁶.

Through mutagenesis screens for touch-insensitive mutants^{1,2}, we previously identified *mec-3*, *mec-17* and eight *mec-3*-dependent genes that are needed for the function of the six touch receptor neurons (refs 7, 8; and G. Gu, L. Emtage and M.C., unpublished data). Those screens identified several alleles for each of these genes (except *mec-17*) and were therefore at or near saturation; however, they would not have identified genes whose activity is redundant, subtle or pleiotropic, or genes whose loss produces touch-supersensitive animals.

Because adult animals contain roughly 3,000 nuclei but only six touch receptor neurons, the identification of differences in RNA from animals with and without these cells on DNA microarrays has an inherent problem of sensitivity. Indeed, we could not identify any

touch receptor genes using RNA from wild-type and *mec-3* mutant animals on DNA microarrays (data not shown). We therefore obtained and cultured wild-type and *mec-3* mutant cells from embryos, isolated touch receptors by cell sorting, and amplified RNA from the isolated cells. We identified the cells by green fluorescent protein (GFP) fluorescence, which was expressed either from the promoter for *mec-18*, a touch-receptor-specific gene, in wild-type embryonic touch receptors (ALML/R and PLML/R), or from the *mec-3* promoter, which is specific to touch receptors and two other neurons (FLPL/R) in the embryo⁴, for *mec-3* mutant cells. Both promoters are expressed after the touch receptors are generated.

When initially dissociated, embryos produced a few faintly fluorescent cells. When cultured overnight, however, more cells (0.3–0.6%) elaborated processes and had intense GFP fluorescence. These GFP-positive cells, which were presumably generated and/or differentiated in culture, were similar in morphology (Fig. 1a) and gene expression (Fig. 1b and Table 1) to *in vivo* touch receptor neurons. The wild-type cultured cells were usually monopolar with a single long process or bipolar with a smaller second process at 180° from the first, in other words, they were similar to *in vivo* ALM or PLM cells, respectively. In contrast, the *mec-3* cells were bipolar or multipolar with smaller processes that branched more randomly.

To characterize gene expression in wild-type and mutant touch receptors, we isolated the cells by size, viability and GFP fluorescence intensity using fluorescence-activated cell sorting. A typical sorting produced 4×10^6 cells, 40–60% of which were GFP positive (~100-fold enrichment). From these cells, we obtained about 100 pg of poly(A)⁺ RNA, which was amplified linearly by roughly one million times. All nine known *mec-3*-dependent *mec* genes were expressed at much higher levels in wild-type cells than in *mec-3* cells (Fig. 1b and Table 1). (*mec-3* is least differentially expressed probably because of its autoregulation; its expression declines in *mec-3* mutants as animals mature⁴; therefore, the reduction of *mec-3* messenger RNA in embryonic cells was probably not complete.) These results show that the differentiated touch receptors in culture express *mec-3*-dependent genes and that linear amplification can successfully amplify the corresponding RNAs.

To identify *mec-3*-dependent genes, we hybridized RNA from sorted wild-type cells and *mec-3* cells to genomic DNA microarrays containing DNA for 17,817 of the 18,967 known or predicted *C. elegans* genes⁹. Three sets of independently prepared RNA samples from mutant and wild-type touch receptors were each hybridized onto two separate arrays. We identified 71 *mec-3*-dependent candi-

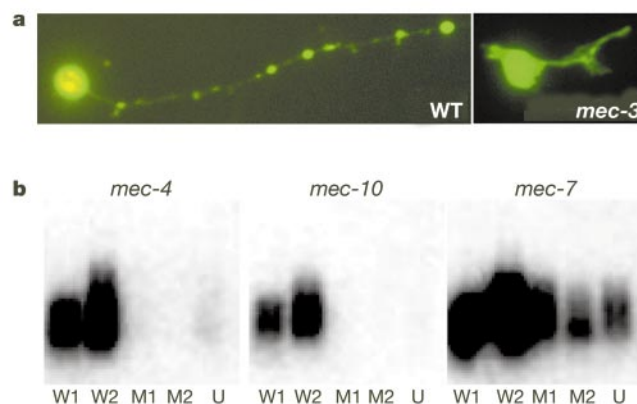


Figure 1 Morphology and gene expression of cultured *C. elegans* cells. **a**, A wild-type (WT) touch receptor neuron and a *mec-3* mutant cell after 12 h in culture. Both cells fluoresce because of expressed GFP. The wild-type cell has a single neurite, whereas the *mec-3* cell has several shorter neurites. **b**, Northern analyses of amplified RNA from sorted wild-type touch receptors (W1 and W2), sorted *mec-3* mutant cells (M1 and M2), and unsorted wild-type cells (U). Samples were probed with cDNAs for the indicated genes.

† Present address: Aventis Pharmaceuticals, 1041 Rt. 202-206, Bridgewater, New Jersey 08807-0800, USA.

date genes (Table 1) by choosing genes whose ratio of fluorescence intensity was ≥ 2.0 and was significant using Student's *t*-test ($P < 0.05$). Using the same criteria, we identified 25 candidate genes that were overexpressed in the *mec-3* mutant cells (data not shown). These latter genes could be either *mec-3*-repressed genes or

genes that are expressed in a *mec-3*-independent manner in the FLP cells. Seven of the nine known *mec-3*-dependent genes and two other touch receptor-expressed genes, *mec-12* (ref. 10) and *unc-24* (T. Barnes and S. Hekimi, personal communication), were among the 71 *mec-3*-dependent candidate genes. Of these 11 genes, 7 were

Table 1 *mec-3*-dependent genes

Gene name*	Description	Array ratio†	PCR ratio†	P	S1‡	S2‡
F57H12.7	MEC-17	62.10	2,928	0.002	•	•
<i>mec-9</i>	EGF/Kunitz repeat protein	12.30	34	0.005		
<i>mec-7</i>	β-Tubulin	11.91	9	0.023	•	•
<i>mec-14</i>	Aldo-keto reductase-like	9.34	20	0.011	•	•
T07D1.3	Aldehyde dehydrogenase domain	8.75	69	0.004	•	•
<i>mec-18</i>	coA synthetase-like	8.32	1,574	0.007	•	•
C03A3.3	Serine protease/metallo- β -lactamase-like	8.13	62	0.018		•
<i>mec-4</i>	Degenerin	7.40	67	0.009	•	•
<i>twk-28</i>	TASK potassium channel subunit	6.42	8	0.003		
<i>mec-12</i>	α-Tubulin	5.91	16	0.001	•	•
F42A9.9	Unknown	4.52	32	0.014	•	
ZC482.5	GABA receptor	4.24	17	0.014		•
F14E5.4	Acid phosphatase	3.91	280	0.005		•
<i>unc-24</i>	Stomatin/lipid transfer protein-like	3.58		0.038		
F20D6.5	Tyrosine kinase catalytic domain	3.33	8	0.002		
C06A8.3	17k antigen (<i>O. volvulus</i>)	3.29	7	0.024		
<i>far-3</i>	<i>O. volvulus</i> antigen peptide-like	3.17	8	0.014		
Y48G8AL.8	Contains ribosomal protein L22 signature	3.13		0.032		•
<i>mec-3</i>	LIM homeodomain transcription factor	3.09	7	0.014	•	•
<i>rpl-18</i>	Ribosomal protein L18	3.05		0.047		
Y54G11A.1	Unknown	2.92		0.029	•	
F01D5.8	Esterase/lipase/thioesterase active site	2.90	24	0.008	•	•
<i>cct-2</i>	T-complex chaperonin protein	2.89	3	0.017	•	
Y39A3CL.1	Unknown	2.81		0.027		•
F29B9.4	Jmjc and weak granin domains	2.79	2	0.004		
D2007.2	Major sperm protein (MSP) domain	2.75		0.026	•	
Y17D7B.4	Unknown	2.74		0.010		•
<i>hsp-16.41</i>	Heat shock protein	2.74		0.040		•
F49E8.4	Cytidine deaminase	2.68	4	0.031		
W05H9.1	Unknown	2.50		0.032	•	
<i>rpl-20</i>	Ribosomal protein L20	2.49	4	0.027		
<i>rps-18</i>	Ribosomal protein S13	2.49	3	0.026		
<i>cct-4</i>	T-complex chaperonin protein	2.47	4	0.021		
W08E12.3	Member of uncharacterized protein family	2.46		0.050	•	•
Y38F2AR.4	Contains similarity to <i>C. elegans</i> MUT-2	2.44		0.046	•	
VW06B3R.1	Insulinase (proteinase M16)	2.42		0.027		
F54E2.2	Unknown	2.33		0.040	•	
W08E12.5	Member of uncharacterized protein family	2.30		0.045	•	
<i>rps-14</i>	Ribosomal protein S14	2.29		0.040	•	
F47B10.7	Acyl-coA binding protein	2.26		0.027	•	
<i>cyp-5</i>	Cyclophilin	2.25		0.047		
C05D11.2	Vacuolar protein sorting protein 16-like	2.21	2	0.001		
<i>cap-1</i>	F-actin capping protein α -subunit	2.21		0.024		•
EGAP7.1	Cuticular collagen	2.18		0.039		
C47D2.1	Unknown	2.17		0.027		
T04B2.3	Unknown; prenyl group binding site	2.16		0.006		
F52E4.1	Propionyl-CoA carboxylase β	2.16		0.029		•
<i>bag-1</i>	Contains ubiquitin domain	2.15		0.023		
F46E10.2	Contains similarity to elastin	2.15		0.003		
F54E12.4	Histone H2B	2.15		0.047		
<i>sra-6</i>	Sra family chemoreceptor	2.14		0.043		
K07C11.4	Similar to type-B carboxylesterase/lipase	2.13	4	0.024		
F19B2.5	Unknown; SNF2_N domain	2.12		0.006		•
<i>cct-1</i>	T-complex chaperonin protein	2.10	4	0.002		
C02F5.3	GTP-binding protein	2.10		0.047		
R05D7.5	Unknown	2.09		0.038		
M03F4.7	Calcium-binding protein	2.09		0.003		
C27B7.6	Serine/threonine protein phosphatase	2.09	4§	0.021		
B0336.11	MSP domain; like VAMP-associated proteins	2.08		0		
F39B2.11	Contains similarity to bacterial GntR	2.08		0.046		
F58E10.4	Unknown; AN1-like zinc finger	2.08		0.038		
F10A3.11	CW domain protein	2.08		0.009	•	•
F46G10.6	HLH DNA-binding domain	2.06		0.009		•
<i>srp-1</i>	Serpin, serine protease inhibitor	2.06		0.005		•
F42H10.3	Unknown; SH3 domain	2.05		0.004		
C55A1.6	Member of uncharacterized protein family	2.03		0.008		
<i>mec-1</i>	EGF/Kunitz repeat protein	2.03	21	0.009	•	•
K03H1.4	Transthyretin-like protein	2.03		0.003	•	•
F49E8.3	Puromycin-sensitive aminopeptidase-like	2.02	3§	0.031		
<i>arf-3</i>	GTP-binding protein	2.01		0.039		•
<i>acr-13</i>	Acetylcholine receptor	2.01		0.013		•

mec genes that have been cloned previously are in bold. *P* values are from the *t*-test.

*Gene names (from the primers that generated the DNA microarray spots) and gene functions are from Wormbase (<http://www.wormbase.org>).

†The expression ratios from microarrays and from quantitative RT-PCR are given.

‡A dot denotes the presence of the MEC-3::UNC-86 consensus binding site (S1) or the heptanucleotide over-represented in the first 20 genes (S2). *mec-10* (PCR ratio 43) and *mec-2* (PCR ratio 70) have both sequences.

§One set of wild-type and mutant RNAs was used for the PCR ratio instead of two.

found among the first 14 genes in Table 1 (an additional *mec* gene, *mec-17*, is also in this group; see below). As described above, *mec-3*, one of the four remaining genes, is probably lower on the list because of the early age of the isolated cells. The remaining genes (*mec-1*, *mec-2* and *mec-10*) were probably lower or are missing from Table 1 because their RNAs hybridized poorly to the arrays. The arrays were spotted with coding-rich sequences from genomic DNA, and the DNAs for these three genes would not hybridize to at least 600 nucleotides at the 3' ends of the RNAs. Our amplification methods, however, preferentially produce sequences from the 3' ends of the RNAs. DNA microarrays that could detect the 3' ends of mRNAs would have more accurately identified the *mec-3*-dependent genes.

Quantitative polymerase chain reaction with reverse transcription (RT-PCR) indicated that the nine known *mec-3*-dependent genes were enriched from 7-fold to 1,574-fold in wild-type cells (Table 1). Eleven newly identified *mec-3*-dependent candidates gave similar ratios and were therefore likely to be expressed primarily in touch receptors and regulated directly by MEC-3. We confirmed this expression for three of these genes (Fig. 3): F57H12.7 and C03A3.3 were expressed in the six touch receptors, whereas T07D1.3 was expressed in touch receptors and the FLP cells. The expression of these genes in the touch receptors required *mec-3*.

Quantitative RT-PCR of 11 other genes in Table 1 gave modest differential ratios (≤ 4). All of these genes ranked below *mec-3*. As described below for the *cct* genes, genes that give small ratios can be expressed differentially in the touch receptors. The modest

difference in expression might result from more general gene expression or an indirect effect on gene expression that results from the differentiation of the wild-type cells. If we assume that the genes above *mec-3* are true positives and use the fraction of previously uncharacterized genes below *mec-3* that have differential expression ratios of ≤ 2 (2/12) to estimate false positives, then 87%

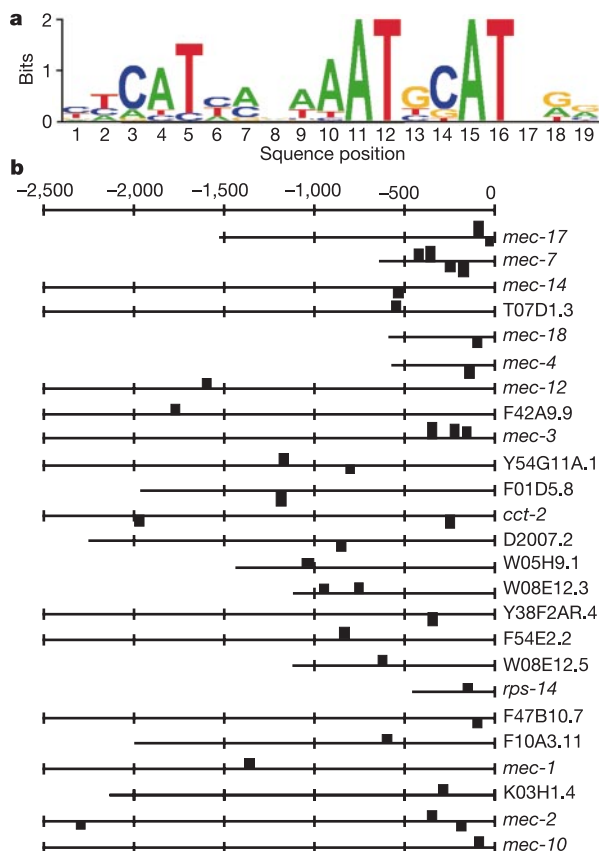


Figure 2 The MEC-3::UNC-86 binding site. **a**, Consensus derived from known MEC-3::UNC-86 binding sites from *mec-3*, *mec-4* and *mec-7* in *C. elegans*, *C. briggsae* and *C. vulgaris*. A perfect match would have a score of two bits. **b**, Positions of significant matches ($P < 10^{-4}$) to the consensus in the regions 5' of the start codons of newly identified and known *mec-3*-dependent genes. The position of the start ATG is taken as 0; the heights of the bars are proportional to the extent that the identified sites match the consensus sequence. Bars shown underneath the lines represent the sequences in the other strand.

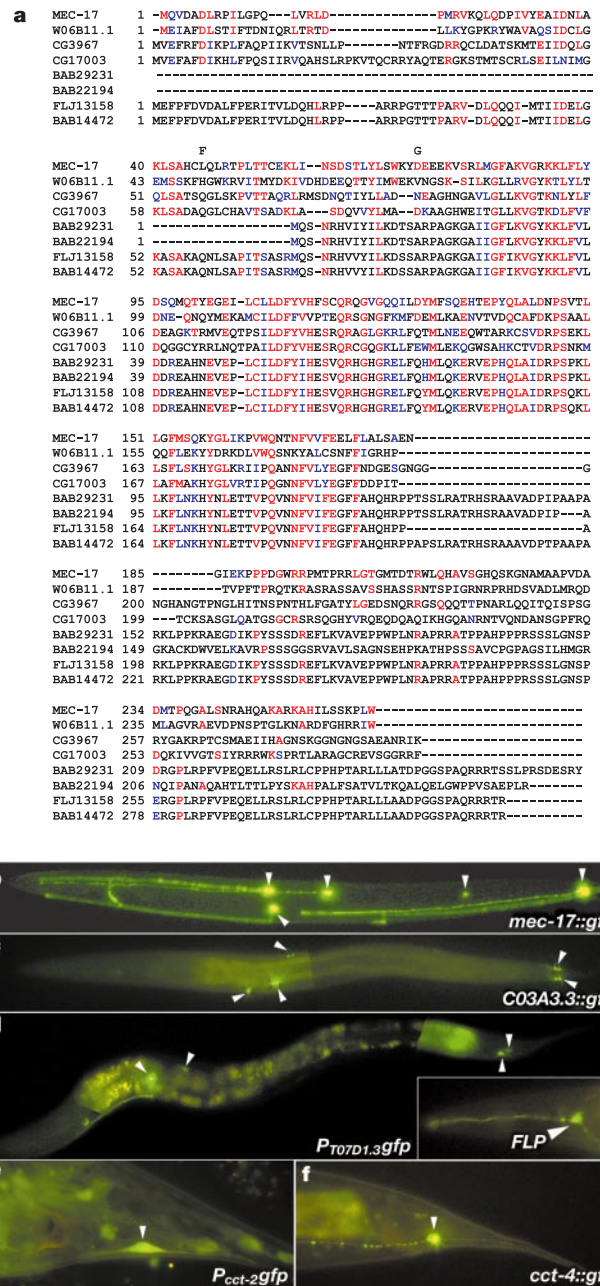


Figure 3 Newly identified *mec-3*-dependent genes. **a**, ClustalW (<http://clustalw.genome.ad.jp>) alignment of MEC-17 and similar proteins from *C. elegans* (W06B11.1), *Drosophila* (CG3967 and CG17003), mouse (BAB29231 and BAB22194) and human (FLJ13158 and BAB14472; BAB names are GenBank accession numbers, others are gene names). Amino acids that are identical (red) or similar (blue) to those in MEC-17 are indicated. The amino acid changes in MEC-17(U265) are indicated and correspond to nucleotide changes CTT (Leu) to TTT (Phe) and GAT (Asp) to GGT (Gly). **b–f**, GFP expression patterns (triangles indicate touch receptor cell bodies) of *mec-17::gfp* in a wild-type L3 larva (**b**), *C03A3.3::gfp* in a wild-type L4 larva (**c**), a *P_{70D1.9}::gfp* promoter fusion in a wild-type adult (**d**), a *P_{cct-2}::gfp* promoter fusion in a wild-type adult PLM touch receptor (**e**) and *cct-4::gfp* in a wild-type adult PLM cell (**f**).

of the genes in Table 1 are likely to be differentially expressed in wild-type touch receptors.

Several factors may contribute to the differences between the array ratios and those obtained by quantitative RT-PCRs. As explained above for *mec-1*, *mec-2* and *mec-10*, some RNAs hybridize poorly to the arrays. In addition, the need to generate a ratio from the array data means that highly expressed RNAs in wild-type cells (that may saturate hybridization) and RNAs such as *mec-4* that are not expressed in *mec-3* mutant cells (which by default have a value that is equal to the standard deviation of the local background) will give lower ratios on the arrays.

MEC-3 maintains its own transcription and activates *mec-4* and *mec-7* by binding to DNA as part of a MEC-3::UNC-86 heterodimer^{6,11,12}. The consensus derived from these binding sites (Fig. 2a) contains the binding site for UNC-86 (AAATT/GCAT) and an additional sequence (TCATCA) positioned 5' to this site. Seventeen newly identified *mec-3*-dependent genes in Table 1 and eight of the nine known *mec-3*-dependent genes contain the consensus sequence in the regions 5' of their start codons (up to 2.5 kb; Fig. 2b). They are presumably candidates for direct targets of *mec-3* regulation.

Analysis of the oligonucleotide frequencies¹³ in the region lying 5' to the start codon of the top 20 genes in Table 1 showed that a heptanucleotide (AATGCAT) is significantly over-represented ($P < 0.01$). This motif matches the UNC-86 binding core in the consensus sequence (Fig. 2a), and, thus provides a proof of principle that binding sites needed for transcription can be deduced from the array data. These results also provide a cautionary note, because this site is needed for UNC-86 binding, not MEC-3 binding. Of the 71 genes in Table 1, 27 contain this heptanucleotide. Consistent with the hypothesis that genes that show strong enrichment are direct targets of *mec-3* regulation, 12 of the 16 genes with an expression ratio of ≥ 9 from quantitative RT-PCR (including *mec-10*) contain this heptanucleotide, whereas none of the 15 genes with an expression ratio of < 9 (excluding *mec-3*, which is a special case because of its autoregulation) has this heptanucleotide. Genes lacking these sites—for example, three of the four genes that encode ribosomal proteins—might be regulated indirectly by differentiation of the touch receptors.

We have noted several general properties of the *mec-3*-dependent genes in Table 1. First, the genes reflect the general distribution of genes on all six chromosomes. Second, this collection of genes is under-represented for genes that are expressed together in operons. Roughly 15% of *C. elegans* genes are initially expressed in multigene transcripts from operons¹⁴. Nine genes known to be members of operons in *C. elegans* are found in Table 1, but the other members of the same operons are not. Either these nine genes are expressed separately from the other members of their operons or they represent false positives. Two genes (F49E8.3 and F49E8.4) are sufficiently close enough to form an operon, but their expression as an operon has not been confirmed (they also have relatively low differential ratios by RT-PCR). Third, even though redundancy owing to gene families may have contributed to genes on this list being missed in our genetic screens, multiple members of only 4 out of 25 gene families are represented by the genes in Table 1. *mec-4* and *mec-10* both encode similar degenerin channel subunits, but these genes do not act redundantly¹⁵. Three genes (*cct-1*, *cct-2* and *cct-4*) encode similar, but probably not redundant components of the T-complex chaperonin (see below). In addition, Table 1 contains *unc-24*, whose product shares a stomatin-like region with MEC-2 (ref. 16), and *acr-13*, an acetylcholine receptor subunit (two other such subunits are expressed in the touch receptors^{17,18} but are not included in Table 1). Thus, Table 1 may contain genes that provide functional redundancy, but redundancy through the use of multiple members of gene families does not seem to have a numerically important role in touch receptor differentiation. Last, the genes used by these cells are not more nematode-specific than are those of the whole *C. elegans* genome. Of the genes in Table 1, 38% (27/71)

show similarity to genes in other organisms, as compared with 36% of all *C. elegans* genes¹⁹.

One motivation for this project was to identify candidates for uncloned, but mutated, genes. The highest ranking gene in Table 1, F57H12.7, is such a gene. F57H12.7 maps to the position of the previously uncloned *mec-17* gene. *mec-17* seems to be needed for maintained differentiation of the touch receptors; *mec-17* mutants are touch sensitive at hatching, but become insensitive as they mature². Expression of *mec-3* and *mec-7* is reduced in older *mec-17* animals^{4,5}. F57H12.7 was identified as *mec-17* because the single mutant allele of *mec-17*, *u265*, contained two missense mutations (Fig. 3a) and microinjection of F57H12.7 genomic DNA into *mec-17(u265)* animals rescued the mutant phenotype. F57H12.7 was expressed solely and strongly in the touch receptors from late embryos to adults in wild-type animals but not in *mec-3* mutants (Fig. 3b and data not shown). *mec-17* encodes a predicted polypeptide of 262 amino acids. Although this protein has not been described previously and has no readily identifiable motifs, MEC-17, another predicted *C. elegans* gene, W06B11.1, and two predicted genes each from *Drosophila*, mouse and humans share a domain of 158 amino acids that has 29–41% identity (Fig. 3a).

Although many of the genes encode proteins of unknown function, others suggest specific roles in the touch receptors. For example, one of the most striking features of the *C. elegans* touch receptors is that their processes are packed with unusual, 15-protofilament microtubules²⁰ that are formed from β -tubulin and α -tubulin encoded by *mec-7* (ref. 7) and *mec-12* (ref. 10), respectively. Thus, the presence of three genes encoding components of the T-complex chaperonin (*cct-1*, *cct-2* and *cct-4*) in Table 1 is apt, because this chaperonin has been implicated in the correct folding of α - and β -tubulin and in microtubule assembly²¹. *cct-2* and *cct-4* were expressed in touch receptors (Fig. 3e, f) in a *mec-3*-dependent fashion (data not shown). These genes were expressed widely in *C. elegans*, but their expression in other cells was not affected by mutation of *mec-3* (data not shown).

The role of many of the genes, especially those with strong *mec-3*-dependent expression, in the function of touch receptors is intriguing. Such genes include *twk-28*, encoding a TASK potassium channel protein; ZC482.5, encoding a GABA (γ -aminobutyric acid) receptor subunit; four genes that encode predicted enzymes or proteins with catalytic domains, C03A3.3 (metallo- β -lactamase), F14E5.4 (acid phosphatase), F20D6.5 (tyrosine kinase) and F01D5.8 (esterase); and two genes (C06A8.3 and *far-3*) that encode proteins that are similar to antigens from the parasitic nematode *Onchocerca volvulus*.

In multicellular organisms, tissue and cell heterogeneity has restricted the usefulness of DNA microarrays²². In *C. elegans*, for example, RNA from whole animals has been applied to DNA microarrays to see gross differences in expression between male and hermaphrodite animals⁹ or germline cells²³, which make up about two-thirds of the cells in adults, or between animals at different developmental stages^{9,24}. But genes that are expressed in a few cells, such as the six touch receptors, cannot be identified systematically using whole-worm RNA. Thus, a main advantage of our technique, which uses isolated cells, linearly amplified RNA, and DNA microarrays, is its ability to analyse global gene expression for individual cell types. Although the cell culturing needed in this method may prevent the detection of differences resulting from cell-cell interactions, these techniques should be applicable towards the study of cellular development and function under many different experimental conditions. □

Methods

Cell culture, FACS and RNA amplification

Strain TU2583, containing a *P_{mec-18gfp}* integrated array, and strain TU2767, containing a *P_{mec-3gfp}* integrated array with *mec-3(e1338)*, were used, respectively, to generate wild-type and *mec-3* mutant touch receptors. Cells were obtained from embryos by a

modification of the method of L. Bloom (<http://cobweb.dartmouth.edu/cgi-bin/cgiwrap/~ambros/protocol.cgi?id=16>). Details of the modified protocol and other methods are given in the Supplementary Information. We carried out all flow cytometric analyses and cell sorting on a FACS VantageSE equipped with CellQuest software (Becton Dickinson). RNA from the sorted cells was amplified through two rounds by a modification of the method of ref. 25.

Real-time RT-PCR

Roughly 3–5 µg of second-round amplified RNA from purified wild-type or *mec-3(e1338)* touch receptors was converted to complementary DNA using random primers and PowerScript reverse transcriptase (Clontech) according to the manufacturer's protocol. We used 30 ng of cDNA from each (two wild-type and two *mec-3(e1338)* RNAs) sample in a 20-µl RT-PCR reaction using the FastStart DNA Master SYBR Green I kit (Roche) supplemented with 3 mM MgCl₂ and Platinum Taq Polymerase (Invitrogen). Quantitative RT-PCR was done using the Roche LightCycler system. The efficiency of amplification for each pair of primers was determined by a standard curve that was generated using serially diluted genomic DNA. The ratio of expression level of each investigated gene between wild-type and *mec-3(e1338)* samples was calculated as described²⁶ using *unc-86* as the reference gene. For at least eight genes, one intron was included in the PCR product from genomic DNA whose length and melting temperature would be different from that of the PCR product from RNA samples. As we did not see the melting curve for the PCR product generated from genomic DNA, we could be certain that there was no detectable genomic DNA contamination in the generated RNA samples.

Microarray analysis

Twenty-microgram samples of amplified RNA were reverse transcribed into cDNAs using random hexamers, labelled with Cy3-dUTP or Cy5-dUTP (Amersham), and applied to microarrays as described²⁵. We scanned hybridized microarrays with an Axon scanner and determined the expression levels with GenePix software. The fluorescence intensities for Cy3-labelled RNA and Cy5-labelled RNA were normalized by GenePix software and the background standard deviations were normalized accordingly. Three amplified RNA samples were generated from three independent primary cell cultures and cell sorting for wild-type touch receptors (w1, w2 and w3) and for *mec-3(e1338)* touch receptors (m1, m2 and m3). Each hybridization (w1–m1, w2–m2, w3–m3) was done twice. The data for these experiments are stored in the Stanford Microarray Database in the *C. elegans* section, under Chalfie (<http://genome-www5.stanford.edu/cgi-bin/SMD/cluster/QuerySetup.pl>). All non-flagged spots for which the net fluorescence intensity at least in one channel was three background standard deviations above background were considered to be well hybridized and measured. If the net fluorescence intensity for a non-flagged spot was within 1 s.d. of the background, the net fluorescence intensity was set as 1 s.d. of the background. We averaged the natural logs of the ratios for the duplicate hybridizations. Student's *t*-test *P* values were determined with the averages from independent experiments.

Gene analysis

We used the motif-finding program MEME²⁷ (<http://meme.sdsc.edu/meme/website/meme.html>) to derive a 19-nucleotide consensus sequence from 19 known MEC-3 UNC-86 binding sites in the promoter region of *mec-3*, *mec-4* and *mec-7* in *C. elegans*, *Caenorhabditis briggsae* and *Caenorhabditis vulgaris*. A weight matrix for the defined consensus site was built on the basis of the multiple sequence alignment and was used to search the regions 5' of the start codon (up to 2.5 kb) of the genes in Table 1 for the significant matches ($P < 10^{-4}$) using a pattern search tool (<http://www.ucmb.ub.ac.be/bioinformatics/rsa-tools>). A graphical representation of the weight matrix was then generated (<http://ep.ebi.ac.uk/EP/SEQLOGO>).

We used the Blast protein–protein tool (BlastP; <http://www.ncbi.nlm.nih.gov/BLAST>) with the non-redundant Protein Data Base at the National Center for Biotechnology Information and with *C. elegans* proteins at Wormbase (<http://www.wormbase.org>) to identify similar genes, that is, those with an expectation value (*E*) $< 10^{-10}$ and that were aligned for at least 80% of their lengths¹⁹.

We amplified the *mec-17(u265)* genomic fragment by PCR from worm lysates and the PCR product was sequenced by GeneWiz Inc. A 5.5-kb fragment of *mec-17* genomic DNA was amplified by PCR using the Clontech polymerase mixture and subcloned using the TA cloning kit (Invitrogen). The plasmid (0.75 ng µl⁻¹) was injected together with the *rol-6(su1006)* marker plasmid pRF4 (75 ng µl⁻¹; ref. 28) and pBR322 (75 ng µl⁻¹) into *mec-17(u265)* animals. The rescue of the touch-insensitive phenotype was determined in transgenic adults.

We generated *mec-17::gfp* by inserting a 5-kb genomic DNA fragment (produced by PCR) including 1.9 kb of upstream sequence and all of the predicted genomic sequence except that for 42 amino acids at the carboxy terminus of MEC-17 into pPD95.77 (a gift from A. Fire), which contains the *gfp* gene with added introns. The construct was transformed with DA735 (ref. 29), a plasmid containing *lin-15* + DNA, into MT1642 (*lin-15(n765)*) to produce three stable lines that were examined using an Axiophot microscope with an FITC filter set. We also transformed TU353 (*mec-3(e1338)*; *dpy-20(e1282)*) with *mec-17::gfp* DNA and pMH86 (ref. 30), a plasmid containing *dpy-20* + DNA, and obtained three stable lines. Animals from these lines were mated with *dpy-20(e1282)* males to produce *mec-3/+* animals.

The *cct-4::gfp* construct was generated by inserting 2,389 bases of *cct-4* genomic DNA, including 585 bases of upstream sequence, in-frame into pPD95.77. *P_{cct-2}gfp* was generated by inserting 2,436 bases of upstream sequence of *cct-2* and 42 bases of coding sequence in-frame into pPD95.77. Each construct was injected with pMH86 into TU353 and produced three stable lines. Animals were mated with *dpy-20(e1282)* males to obtain *mec-3/+* heterozygotes.

Similar constructs were formed for T07D1.3 by inserting a 2.2-kb DNA fragment including the upstream sequence and the starting codon of the gene into pPD95.75 (a gift from A. Fire), and for C03A3.3 by inserting a 3.6-kb genomic DNA fragment including the upstream sequence and all of the predicted genomic sequence for the gene in-frame into pPD95.77.

Received 5 February; accepted 20 May 2002; doi:10.1038/nature00891.

- Chalfie, M. & Sulston, J. Developmental genetics of the mechanosensory neurons of *Caenorhabditis elegans*. *Dev. Biol.* **82**, 358–370 (1981).
- Chalfie, M. & Au, M. Genetic control of differentiation of the *Caenorhabditis elegans* touch receptor neurons. *Science* **243**, 1027–1033 (1989).
- Way, J. C. & Chalfie, M. *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell* **54**, 5–16 (1988).
- Way, J. C. & Chalfie, M. The *mec-3* gene of *Caenorhabditis elegans* requires its own product for maintained expression and is expressed in three neuronal cell types. *Genes Dev.* **3**, 1823–1833 (1989).
- Mitani, S., Du, H., Hall, D. H., Driscoll, M. & Chalfie, M. Combinatorial control of touch receptor neuron expression in *Caenorhabditis elegans*. *Development* **119**, 773–783 (1993).
- Xue, D., Finney, M., Ruvkun, G. & Chalfie, M. Regulation of the *mec-3* gene by the *C. elegans* homeoproteins UNC-86 and MEC-3. *EMBO J.* **11**, 4969–4979 (1992).
- Savage, C. *et al.* *mec-7* is a β -tubulin gene required for the production of 15- protofilament microtubules in *Caenorhabditis elegans*. *Genes Dev.* **3**, 870–881 (1989).
- Gu, G., Caldwell, G. A. & Chalfie, M. Genetic interactions affecting touch sensitivity in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **93**, 6577–6582 (1996).
- Jiang, M. *et al.* Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **98**, 218–223 (2001).
- Fukushige, T. *et al.* MEC-12, an α -tubulin required for touch sensitivity in *C. elegans*. *J. Cell Sci.* **112**, 395–403 (1999).
- Xue, D., Tu, Y. & Chalfie, M. Cooperative interactions between the *Caenorhabditis elegans* homeoproteins UNC-86 and MEC-3. *Science* **261**, 1324–1328 (1993).
- Duggan, A., Ma, C. & Chalfie, M. Regulation of touch receptor differentiation by the *Caenorhabditis elegans* *mec-3* and *unc-86* genes. *Development* **125**, 4107–4119 (1998).
- Van Helden, J., Andre, B. & Collado-Vides, J. A website for the computational analysis of yeast regulatory sequences. *Yeast* **16**, 177–187 (2000).
- Blumenthal, T. *et al.* A global analysis of *Caenorhabditis elegans* operons. *Nature* **417**, 851–854 (2002).
- Huang, M. & Chalfie, M. Gene interactions affecting mechanosensory transduction in *Caenorhabditis elegans*. *Nature* **367**, 467–470 (1994).
- Barnes, T. M., Jin, Y., Horvitz, H. R., Ruvkun, G. & Hekimi, S. The *Caenorhabditis elegans* behavioural gene *unc-24* encodes a novel bipartite protein similar to both erythrocyte band 7.2 (stomatins) and nonspecific lipid transfer protein. *J. Neurochem.* **67**, 46–57 (1996).
- Treinin, M. & Chalfie, M. A mutated acetylcholine receptor subunit causes neuronal degeneration in *C. elegans*. *Neuron* **14**, 871–877 (1995).
- Treinin, M., Gillo, B., Liebman, L. & Chalfie, M. Two functionally dependent acetylcholine subunits are encoded in a single *Caenorhabditis elegans* operon. *Proc. Natl Acad. Sci. USA* **95**, 15492–15495 (1998).
- Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215 (2000).
- Chalfie, M. & Thomson, J. N. Structural and functional diversity in the neuronal microtubules of *Caenorhabditis elegans*. *J. Cell Biol.* **93**, 15–23 (1982).
- Liang, P. & MacRae, T. H. Molecular chaperones and the cytoskeleton. *J. Cell Sci.* **110**, 1431–1440 (1997).
- Lockhart, D. J. & Winzler, E. A. Genomics, gene expression and DNA arrays. *Nature* **405**, 827–836 (2000).
- Reinke, V. *et al.* A global profile of germline gene expression in *C. elegans*. *Mol. Cell* **6**, 605–616 (2000).
- Hill, A. A., Hunter, C. P., Tsung, B. T., Tucker-Kellogg, G. & Brown, E. L. Genomic analysis of gene expression in *C. elegans*. *Science* **290**, 809–812 (2000).
- Eberwine, J. *et al.* Analysis of gene expression in single live neurons. *Proc. Natl Acad. Sci. USA* **89**, 3010–3014 (1992).
- Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, E45–E45 (2001).
- Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
- Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
- Huang, L. S., Tzou, P. & Sternberg, P. W. The *lin-15* locus encodes two negative regulators of *Caenorhabditis elegans* vulval development. *Mol. Biol. Cell* **5**, 395–411 (1994).
- Han, M. & Sternberg, P. W. Analysis of dominant-negative mutations of the *Caenorhabditis elegans* *let-60* ras gene. *Genes Dev.* **5**, 2188–2198 (1991).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank X. Chen for help with analysis of the microarray data; M. Goodman for constructing the integrated *mec-18::gfp* strain; B. Tycko for the LightCycler; and E. Schwarz, J. Wang and J. Eberwine for discussion. This work was supported by a grant from the National Center for Research Resources to S.K.K. and a grant from the National Institutes of Health to M.C.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.C. (e-mail: mc21@columbia.edu).

Synaptotagmins I and IV promote transmitter release independently of Ca²⁺ binding in the C₂A domain

Iain M. Robinson^{*,†}, Ravi Ranjan^{†,‡} & Thomas L. Schwarz[‡]

^{*} Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

[‡] Division of Neuroscience, Children's Hospital, Boston, and Department of Neurobiology, Harvard Medical School, 300 Longwood Avenue, Boston, Massachusetts 02115, USA

[†] These authors contributed equally to this work

At nerve terminals, a focal and transient increase in intracellular Ca²⁺ triggers the fusion of neurotransmitter-filled vesicles with the plasma membrane. The most extensively studied candidate for the Ca²⁺-sensing trigger is synaptotagmin I, whose Ca²⁺-dependent interactions with acidic phospholipids and syntaxin¹ have largely been ascribed to its C₂A domain^{2–6}, although the C₂B domain also binds Ca²⁺ (refs 7, 8). Genetic tests of synaptotagmin I have been equivocal as to whether it is the Ca²⁺-sensing trigger of fusion^{6,9–15}. Synaptotagmin IV, a related isoform that does not bind Ca²⁺ in the C₂A domain, might be an inhibitor of release^{16,17}. We mutated an essential aspartate of the Ca²⁺-binding site of the synaptotagmin I C₂A domain and expressed it in *Drosophila* lacking synaptotagmin I. Here we show that, despite the disruption of the binding site, the Ca²⁺-dependent properties of transmission were not altered. Similarly, we found that synaptotagmin IV could substitute for synaptotagmin I. We conclude that the C₂A domain of synaptotagmin is not required for Ca²⁺-dependent synaptic transmission, and that synaptotagmin IV promotes rather than inhibits transmission.

Drosophila melanogaster synaptotagmin I is homologous and highly similar to the mammalian synaptotagmins, and possesses similar biochemical properties (Fig. 1). Aspartate residues 223, 229, 282, 284 and 290 in *Drosophila* (hereafter D1–D5) correspond to those that coordinate Ca²⁺ binding in rat synaptotagmin I (ref. 18). D2 and D3 form the high-affinity Ca²⁺ binding site, and mutations of D2 and D3 are the most deleterious to the Ca²⁺-dependent interactions of the C₂A domain. Mutating D2 or D3 to asparagine abolishes Ca²⁺-dependent binding to syntaxin and phospholipids^{18–20}. Furthermore, the synaptotagmin IV isoform has a serine in the D3 position and consequently does not bind either phospholipids or syntaxin in a Ca²⁺-dependent manner^{16,21}.

To test the functional significance of Ca²⁺ binding to the C₂A domain, we substituted an asparagine for the D2 residue and assayed the Ca²⁺ dependence of syntaxin binding (Fig. 1c). In the absence of Ca²⁺, little binding was observed of syntaxin to a fusion protein of glutathione S-transferase (GST) and the C₂A domain; in 1 mM Ca²⁺, binding was greatly enhanced. The D2N mutation prevented this binding even in the presence of Ca²⁺; the syntaxin bound by the mutated C₂A was similar to that of GST alone. When the synaptotagmin construct included C₂A and C₂B, the syntaxin–synaptotagmin I interaction was independent of Ca²⁺, as has been noted by others⁴, and persisted in synaptotagmin I D2N (data not shown). Binding of phospholipid to the wild-type *Drosophila* C₂A domain was dependent on Ca²⁺ and was abrogated by the D2N mutation (Fig. 1d). The D2N mutation reduced, but did not entirely block, Ca²⁺-dependent binding of phospholipids to a synaptotagmin I construct containing both C₂ domains (Fig. 1e). Thus, mutating this aspartate drastically altered the Ca²⁺-dependent interactions of the C₂A domain of synaptotagmin I, and revealed a residual Ca²⁺ dependence on some interactions that is likely to reside in the C₂B domain⁸.

Although the equivalence of the *Drosophila* and mammalian

synaptotagmins could not be presumed from sequence alone, the D2 residue of the *Drosophila* gene is necessary for conferring Ca²⁺ sensitivity on protein–protein and protein–phospholipid interactions of the C₂A domain of synaptotagmin I. The use of this mutation here does not depend on the total ablation of all Ca²⁺ binding to this domain. It rests instead on the biochemical and structural evidence that the mutation causes substantial changes to the site. If the C₂A domain binds the Ca²⁺ ions that trigger membrane fusion, and if these protein–protein or protein–lipid interactions are related to the initiation of fusion, a dramatic alteration in Ca²⁺ affinity should be mirrored in a profound alteration in the ability of Ca²⁺ to stimulate transmitter release.

To test the functional competence of synaptotagmin I^{D2N}, it and wild-type synaptotagmin I one were epitope tagged, placed under the control of an upstream activating sequence (UAS) promoter, and reintroduced into the fly (*P[UAS-HA-syt]* and *P[UAS-D2N]*). The expression of each construct was driven by the heat-shock-dependent driver *hsp70-Gal4* (ref. 15). This expression system approximately doubled the amount of synaptotagmin I protein present in the adult head compared with *synaptotagmin I (sytl)* heterozygotes (see Supplementary Information), although some of

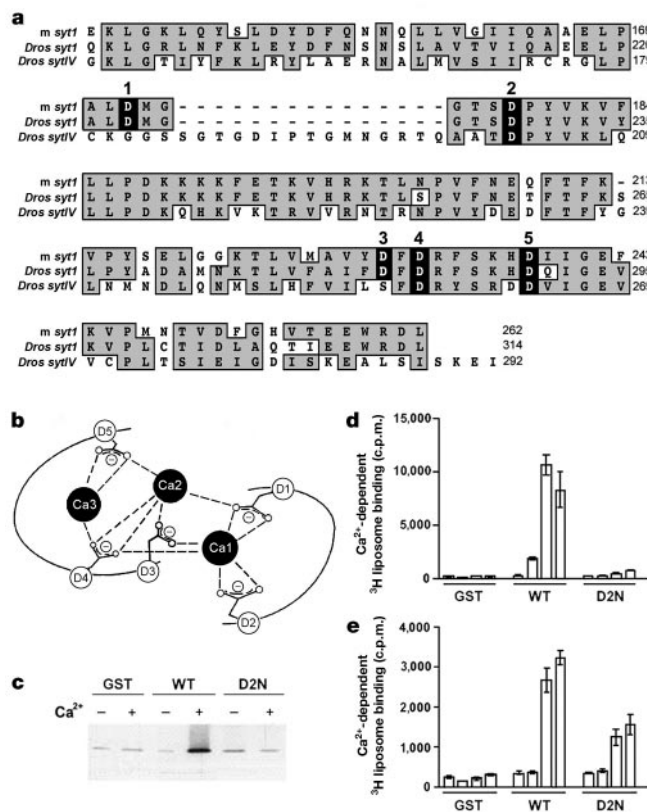


Figure 1 Altered Ca²⁺-dependent interactions of the C₂A domain of synaptotagmin I. **a**, Rat (m) synaptotagmin I C₂A aligned against *Drosophila* synaptotagmins I and IV. Aspartates that coordinate Ca²⁺ ions are numbered D1–D5. **b**, Ca²⁺ binding in the C₂A domain (after ref. 18). **c**, Interaction of *Drosophila* synaptotagmin I C₂A with syntaxin (residues 4–269) in the presence (+) or absence (–) of 1 mM Ca²⁺. Wild-type (WT) and mutant (D2N) GST-fusion proteins, or GST alone, were immobilized on glutathione–agarose beads. Ca²⁺-dependent binding was observed for wild-type C₂A but binding to D2N was no stronger than to GST alone (*n* = 6). **d**, The Ca²⁺-dependent binding of ³H-liposomes to C₂A–GST constructs inhibited by the D2N mutation. **e**, The D2N mutation also reduced Ca²⁺-dependent binding of ³H-liposomes to C₂A–C₂B fusion proteins, although not to background levels. Bars represent (left to right) 0.8, 200 and 2,000 μM free Ca²⁺ in **d** and 0, 0.4, 8 and 2,000 μM in **e**. Each assay point was performed in triplicate and the data shown are representative of six repeats. **d**, **e**, Error bars are standard error of the mean.

this protein may reside in non-neuronal tissues. Each synaptotagmin transgene and driver were crossed into a *sytI*-null background, *syt^{AD4}/In(2)^{D27}* (ref. 11). The *sytI* gene encodes the major synaptotagmin isoform at neuromuscular junctions and its absence greatly reduces transmitter release¹⁰. Thus these null genotypes provide a suitable background for judging the efficacy of a synaptotagmin in which Ca^{2+} binding has been altered.

Both control and D2N transgenes rescued the synaptic phenotype in a heat-shock-dependent manner (Fig. 2a, b). The amplitude of excitatory junctional potentials (EJPs) in the *P[UAS-HA-syt]* larvae fell within the range observed in the *syt⁺* control (*cn,bw*), although they were on average smaller (40.4 ± 7.3 mV, mean $\pm$ s.e.m., $n = 5$ for control; 29.5 ± 3.7 mV, $n = 5$ for *P[UAS-HA-syt]*), suggesting that the heat-shock-driven wild-type transgene was not a perfect substitute for the normal gene. The *P[UAS-D2N]* transgene also had robust EJPs (18.9 ± 4.8 mV, $n = 5$). In a second set of experiments, an *elav-Gal4* driver was used to induce neuronal expression of *P[UAS-D2N]*; under this circumstance EJP amplitudes were 33.8 ± 1.1 mV, compared with 47.1 ± 1.6 mV for *w¹¹¹⁸*, the *syt^{+/+}* control. These responses were fast, tightly coupled to nerve

stimulation, and very similar in time course to those of the wild type. Thus, despite mutation of the essential aspartate, the transgene can alleviate most of the disruption of synaptic transmission caused by removing synaptotagmin I.

The Ca^{2+} dependence of transmission mediated by the synaptotagmin I^{D2N} construct was examined (Fig. 2c) by looking at the apparent cooperativity (N) of release from the slope of double-log plots of quantal content versus extracellular Ca^{2+} concentration. Control *cn,bw* larvae had a Ca^{2+} dependence of neurotransmitter release of $N = 4.0 \pm 0.3$, as previously reported¹². *P[UAS-HA-syt]* larvae had a Ca^{2+} dependence of $N = 3.2 \pm 0.6$ and *P[UAS-D2N]* larvae had a dependence of $N = 2.9 \pm 0.1$. These values are not

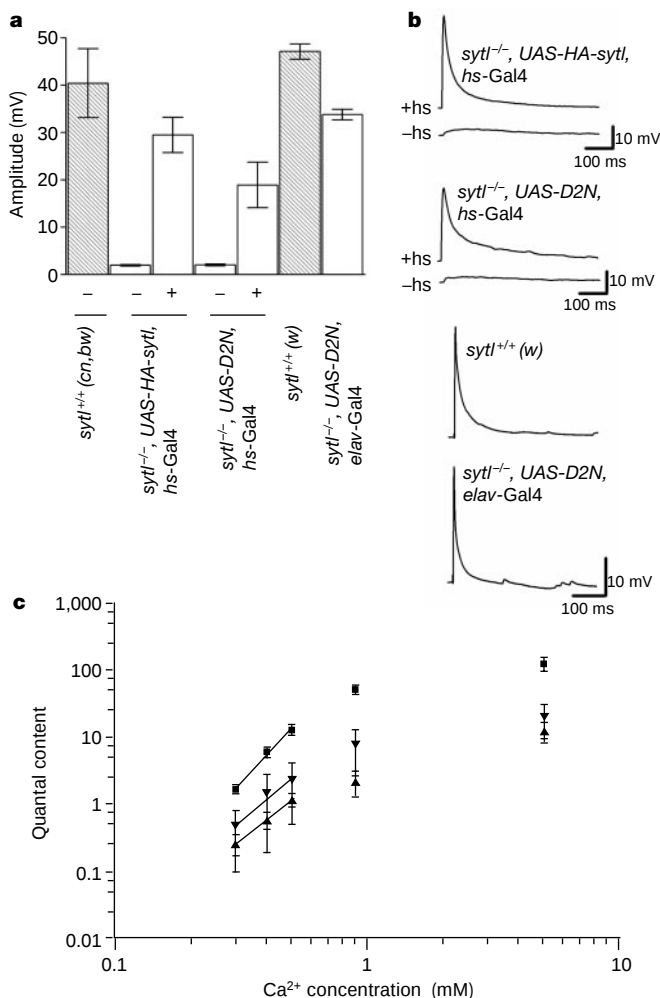


Figure 2 Synaptotagmin I^{D2N} rescues transmission in *sytI*-null mutants. **a**, Rescue of synaptic function by *UAS-sytI^{D2N}* under control of the heatshock-activated *hsp70-Gal4* or *elav-Gal4* driver. Mean $\pm$ s.e.m. amplitudes for EJPs from larvae expressing wild-type (HA-sytI) or D2N mutant synaptotagmin I are compared with *syt^{+/+}* larvae (*cn,bw* or *w¹¹¹⁸*). ($n = 5$) **b**, Representative EJPs from genotypes in **a**. **c**, Ca^{2+} dependence of synaptic transmission is unaffected by the D2N mutation. A double-log plot of mean quantal content versus extracellular $[\text{Ca}^{2+}]$ for control (*cn,bw*; $n = 6$, squares) larvae and *syt*-null larvae rescued with heat-shock-driven *UAS-HA-syt* ($n = 9$) (inverted triangles) or *UAS-D2N* ($n = 10$) (triangles).

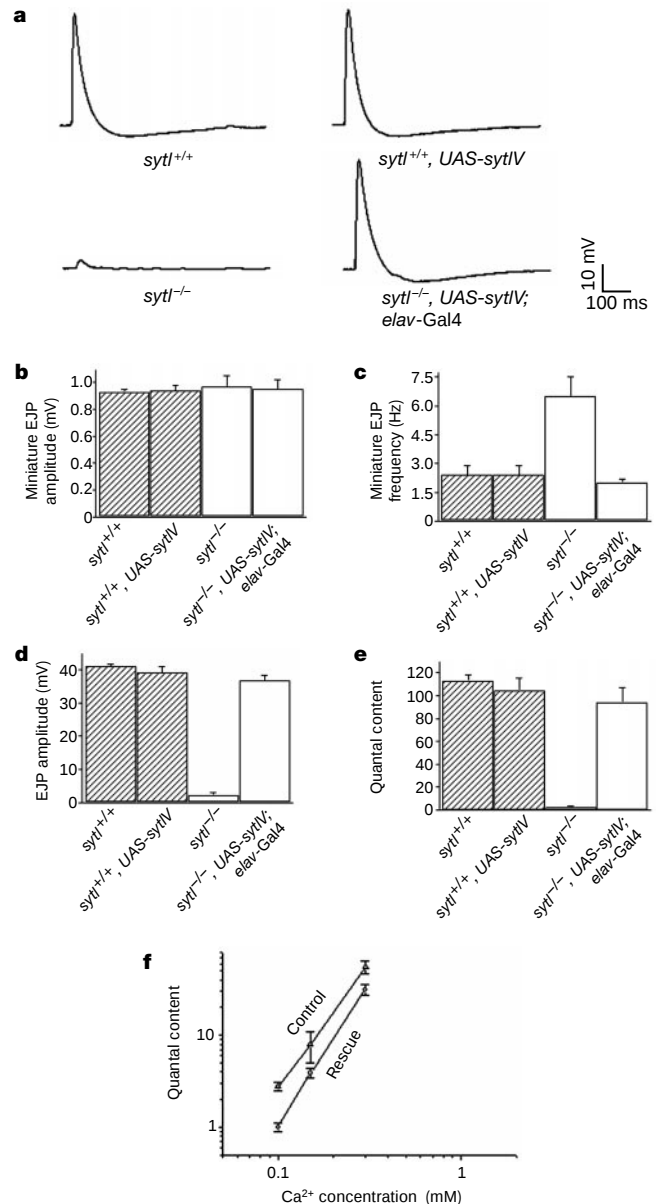


Figure 3 Synaptotagmin IV expression restores transmission to a *sytI*-null synapse. **a**, Representative EJPs from a *syt^{+/+}* control (*w*, $n = 5$), a *syt^{+/+}* control bearing the UAS transgene, but no activator (*UAS-sytIV, syt^{+/+}*, $n = 5$), the null *syt^{-/-}* (*syt^{AD4}/syt^{AD4}*, $n = 6$), and the *syt^{-/-}* mutant rescued with the neuronally expressed synaptotagmin IV transgene (*syt^{-/-}, UAS-sytIV, elav-Gal4*, $n = 11$). **b–e**, For the genotypes in **a**, the mean $\pm$ s.e.m. of the miniature EJP amplitude (**b**), miniature EJP frequency (**c**), EJP amplitude (**d**), and quantal content (**e**). Control lines are hatched. **f**, A double-log plot of quantal content versus extracellular $[\text{Ca}^{2+}]$ for *w⁻* (control) and *syt^{AD4/AD4}, UAS-sytIV, elav-Gal4* ('rescue') genotypes.

significantly different ($P > 0.05$ by pairwise analysis of variance). In the case of both UAS rescue lines, the value for N is likely to be an underestimate because the sample population included some larvae with weak rescue whose evoked responses were difficult to discriminate from spontaneous release, which tended to flatten the dose-response curves. In larvae with robust rescue, examples with slopes of four or greater were observed in both D2N and control rescue lines. Thus a major alteration of the C₂A Ca²⁺-binding site had a negligible effect on the Ca²⁺-dependent properties of the synapse. This result stands in contrast to a previous study that looked at minor perturbations of Ca²⁺ binding in the C₂A domain and observed a decrease in release probability that was interpreted as a correlation of those changes⁶.

These observations prompted us to reinvestigate the function of synaptotagmin IV, the isoform whose sequence contains a serine substitution for the D3 aspartate and whose inability to bind Ca²⁺ in the C₂A domain causes it to act as an inhibitor of release¹⁶. To this end, the UAS-*sytlIV* transgene (line 1,024; ref. 16), with *elav*-Gal4 to activate neuronal expression, was crossed into the *sytl*-null background (Fig. 3a). Larvae lacking synaptotagmin I typically die at

early stages, but, if separated from their heterozygous siblings, they can survive to adulthood²². Nonetheless, transmission at this synapse is nearly absent: in 1 mM Ca²⁺, the EJP in homozygous *sytl*^{AD4} larvae was 2.3 ± 0.7 mV compared with 41.1 ± 0.7 mV in *w*; *sytl*^{+/+} controls or 39.1 ± 2.0 mV in *sytl*^{+/+} controls carrying the UAS-*sytlIV* gene with no driver. When *sytlIV* was expressed in the *sytl*^{AD4/AD4} background, transmission was restored to near-control levels (35.7 ± 1.6 mV; Fig. 3d, e). Consistent with previous studies^{10,11}, we observed a modest increase in the frequency, but not the amplitude, of spontaneous miniature EJPs in *sytl*-null larvae. This phenomenon was reversed by the expression of synaptotagmin IV (Fig. 3b). Thus, despite the inability of the C₂A domain of synaptotagmin IV to bind Ca²⁺ and the reported interference of this protein with Ca²⁺-dependent biochemical processes¹⁶, synaptotagmin IV substitutes efficiently for synaptotagmin I in synaptic transmission.

We attempted, therefore, to reproduce the reported inhibitory effect of synaptotagmin IV expression in a wild-type background, using the same UAS-*sytlIV* construct and *elav*-Gal4 driver¹⁶. We found no effect of synaptotagmin IV on EJP amplitude, quantal content, miniature frequency or miniature amplitude (Fig. 4). Nor was any effect observed when two other *elav*-Gal4 drivers were used (Fig. 4). Comparisons were made at 0.3, 1 and 3 mM Ca²⁺ (data not shown) with the same result. We conclude that overexpression of synaptotagmin IV does not downregulate transmission and that synaptotagmin IV, like synaptotagmin I, promotes transmitter release.

The synaptotagmin I^{D2N} mutation afforded a stringent test of the hypothesis that the C₂A domain of synaptotagmin I is the Ca²⁺ sensor that transduces the change in cytosolic Ca²⁺ into a signal for exocytosis. This transgene restored fast, Ca²⁺-dependent transmission with properties extremely similar to those of wild-type animals and *sytl*-null animals that had been rescued with a wild-type transgene. Similarly, although synaptotagmin IV does not bind Ca²⁺ in its C₂A domain, it efficiently supported transmission in a *sytl*-null background. These findings are not easily reconciled with the hypothesis that the C₂A domain of synaptotagmin is the Ca²⁺-sensing trigger. The release observed was dependent on the induction of the transgenes, so transmission in these animals should predominantly reflect the properties of the transgene and not those of any minor second synaptotagmin gene that might be present. If the mutated Ca²⁺-binding site was indeed the trigger for fusion, alterations in the Ca²⁺ dependence of transmission should have been observed. We cannot rule out that some residual Ca²⁺ binding takes place in the D2N mutation, although biochemical assays indicate it must be very slight (Fig. 1 and refs 18–20), but the profound reduction in Ca²⁺ binding that occurred should have been mirrored by a sharp decrease in synaptic function.

The apparent cooperativity of Ca²⁺ action in the nerve terminal is often used to characterize the sensor. This cooperativity has been inferred from the non-linearity of secretion to changes in extracellular Ca²⁺, which should, at low Ca²⁺ concentrations, be proportional to local change in cytosolic Ca²⁺. The ability of synaptotagmin to bind multiple Ca²⁺ ions has been put forward as a potential correlate for the multiple predicted Ca²⁺ ions that trigger exocytosis^{18,20,23}. This parameter (determined from the slope of log-log plots like that in Figs 2c and 3f) should be a sensitive reporter of changes in the triggering site, including changes in ions bound per site or the removal of one of several Ca²⁺-binding steps. Thus, the observation that this property was indifferent to the D2N mutation is a further indication that the C₂A domain is not involved in the triggering of exocytosis. Similarly, the range of Ca²⁺ concentrations that evoked release was unaffected, suggesting that the Michaelis constant (K_m) for the sensor had not been greatly altered.

Our finding that Ca²⁺ binding by the C₂A domain is not essential for release raises the possibility that the C₂B domain fulfils this function or that the two domains share the task. Yet, if Ca²⁺ sensing

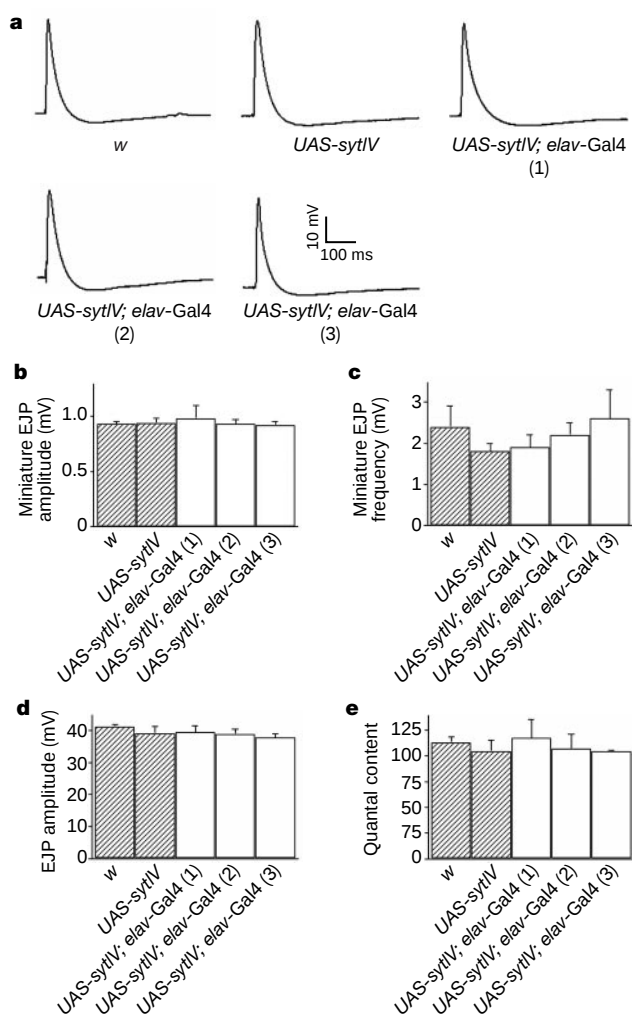


Figure 4 Expression of synaptotagmin IV at wild-type synapses (*sytl*^{+/+}) does not suppress transmission. **a**, Representative EJPs from two control lines (*w*, and *UAS-sytlIV* in which the transgene is present but not activated; $n = 5$ for each) and from three lines that overexpress synaptotagmin IV transgene (*UAS-sytlIV*) in neurons owing to one of three *elav*-Gal4 drivers (**a**, $n = 5$; **b**, $n = 6$; **c**, $n = 3$). **b–e**, For the lines in **a**, the mean $\pm$ s.e.m. of the miniature EJP amplitude (**b**), miniature EJP frequency (**c**), EJP amplitude (**d**) and quantal content (**e**). Control lines without transgenic synaptotagmin IV expression are hatched.

for fusion were accomplished by Ca^{2+} binding to both C_2A and C_2B , then the mutations in our study should have affected transmission significantly, and the apparent cooperativity of fusion should have been reduced by the impairment of C_2A function. Functions of C_2B that are independent of C_2A should be addressed directly with further mutations (see ref. 24).

In *Drosophila*, *Caenorhabditis elegans* and mice, some synaptic transmission persisted in the absence of synaptotagmin I, and this transmission was dependent on Ca^{2+} concentration similarly to the wild type^{9–11,13}. Our present observation that synaptotagmin IV can participate in fast, Ca^{2+} -dependent transmission suggests that synaptotagmin IV or another isoform could mediate the residual release in *sytl*-null mutants. In contrast, that Ca^{2+} -dependent properties of transmission are constant in the face of the different Ca^{2+} -binding properties of either synaptotagmin IV or I^{D2N} suggests that Ca^{2+} binding, at least by the C_2A domain, is not the primary means by which synaptotagmin promotes fusion. Decreased release, which was consistently observed in synaptotagmin I mutations, could instead be accounted for by a defect in any of several steps in nerve terminal function, including a decrease in the pool of releasable docked vesicles or a decrease in the probability of fusion of a docked vesicle^{11,25,26}. This model would be consistent with the interactions of synaptotagmin with plasma membrane proteins and the AP-2 complex^{27,28}, and with ultrastructural studies of mutants^{14,15}. The Ca^{2+} -binding site in the C_2 domains of synaptotagmins may permit Ca^{2+} to modulate vesicle docking or endocytosis, and thereby have subtler synaptic effects that were not assayed here.

Our study reopens the question of the nature of the Ca^{2+} sensor that triggers release. In addition to the C_2B domain of synaptotagmin, many other candidates have been identified²⁹, and multiple sensors may exist to account for the steep dependence of fusion on Ca^{2+} concentration. These alternatives now merit closer examination. □

Methods

Rescue constructs

A haemagglutinin (HA) epitope was incorporated into the intravesicular tail of *Drosophila* synaptotagmin I at the unique *Bst*EII site of the complementary DNA³⁰ by ligating in a pair of complementary oligonucleotides encoding the peptide YPYDVPDYA and selecting a clone (called *HA-syt*) with three tandem repeats of the epitope. The C_2A domains of *Drosophila* and rat synaptotagmins were aligned with the Clustal W algorithms in MacVector version 6.5 (Oxford Molecular). Site-directed mutation of aspartate 229 to asparagines (D2N) was generated by a strategy based on polymerase chain reaction (PCR). The PCR product was subcloned into the *HA-syt* construct and placed in the pUAST vector.

In vitro binding assays

Fusion proteins consisted of GST linked by a thrombin-cleavage sequence to either the C_2A domain of wild-type or D2N synaptotagmin (amino acids 191–319) or syntaxin (amino acids 4–269) (see Supplementary Information). Their binding to syntaxin was assayed according to ref. 4. Lipid binding was assayed by the binding of liposomes to synaptotagmin I–GST fusion proteins³. Details of these assays appear in the Supplementary Information.

Fly strains

After P-element-mediated transformation with the rescue constructs, balanced stocks of transformed animals were crossed to *In(2)syt^{D27}* or *syt^{AD411}*. The second chromosome was balanced with either *CyO_y⁺* or *CyO_{GFP}* so that *syt* homozygotes could be identified by their γ^- or GFP⁺ phenotype. For flies transformed with pUAST vectors, *syt^{AD4}* was recombined with an *hs-Gal4* P-element insert on the second chromosome¹⁵.

Electrophysiology

Third-instar larvae at the 'wandering' stage were dissected and used for electrophysiology in 70 mM NaCl, 5 mM KCl, 5 mM CaCl_2 , 2 mM MgCl_2 , 10 mM NaHCO_3 , 5 mM trehalose, 158 mM sucrose and 5 mM HEPES buffer at pH 7.2 for all recordings, except those in which the Ca^{2+} dependence of transmitter release was determined (see below). *UAS-HA-sytI* or *UAS-D2N* transformants under the control of an *hsp70-Gal4* promoter were induced by heat shocking for 30 min at 37°C every 12 h after egg deposition. Control larvae (no heat shock) were raised at either 18°C (*UAS-HA-sytI*) or 25°C (*UAS-D2N*), and in each case synaptic transmission was substantially weaker than after heat shock. *elav-Gal4* promoter stocks and their controls were all raised at 25°C. All recordings were made in muscle cell 6, abdominal segment⁴¹¹. Data were acquired by an Axoclamp 2-A amplifier

(Axon Instruments) in bridge mode, filtered at 1 kHz, and digitized at 10 kHz before being stored and analysed with pClamp versions 6.0.2 and 8.0.2 (Axon Instruments). Histograms were constructed by averaging 50 separate events from an individual muscle cell stimulated at 1 Hz (Fig. 2) or 15 events at 0.2 Hz (Figs 3 and 4) and then calculating the mean response from at least five cells per line.

The Ca^{2+} dependence of neurotransmitter release was determined in solutions containing 100 μM to 5 mM CaCl_2 , added to 70 mM NaCl, 5 mM KCl, 10 mM MgCl_2 , 10 mM NaHCO_3 , 5 mM trehalose, 136 mM sucrose and 5 mM HEPES buffer at pH 7.2. To prevent low Ca^{2+} concentrations from generating spontaneous action potentials and nerve hyperexcitability, 10 mM MgCl_2 was necessary in this solution. In each Ca^{2+} concentration, 50 stimuli were given at 1 Hz (*UAS-D2N*) or 15 stimuli at 0.2 Hz (*UAS-sytIV*). The responses were averaged for an individual larva and each data point thus obtained was averaged with similar data from 6–10 different animals. Data were corrected for nonlinear summation³¹ and an estimate of the contribution from spontaneous events was subtracted. Spontaneous events were estimated from recordings in the absence of stimulation; events that occurred in the time frame used to measure evoked responses were averaged. Quantal content was plotted against Ca^{2+} concentration and the slope of the curve was used to determine the Ca^{2+} dependence of release.

Received 22 April; accepted 17 June 2002; doi:10.1038/nature00915.

Published online 7 July 2002.

- Chapman, E. R., Hanson, P. I., An, S. & Jahn, R. Ca^{2+} regulates the interaction between synaptotagmin and syntaxin 1. *J. Biol. Chem.* **270**, 23667–23671 (1995).
- Chapman, E. R. & Jahn, R. Calcium-dependent interaction of the cytoplasmic region of synaptotagmin with membranes. Autonomous function of a single C_2 -homologous domain. *J. Biol. Chem.* **269**, 5735–5741 (1994).
- Davletov, B. A. & Sudhof, T. C. A single C_2 domain from synaptotagmin I is sufficient for high affinity Ca^{2+} /phospholipid binding. *J. Biol. Chem.* **268**, 26386–26390 (1993).
- Kee, Y. & Scheller, R. H. Localization of synaptotagmin-binding domains on syntaxin. *J. Neurosci.* **16**, 1975–1981 (1996).
- Li, C., Davletov, B. A. & Sudhof, T. C. Distinct Ca^{2+} and Sr^{2+} binding properties of synaptotagmins. Definition of candidate Ca^{2+} sensors for the fast and slow components of neurotransmitter release. *J. Biol. Chem.* **270**, 24898–24902 (1995).
- Fernandez-Chacon, R. et al. Synaptotagmin I functions as a calcium regulator of release probability. *Nature* **410**, 41–49 (2001).
- Damer, C. K. & Creutz, C. E. Synergistic membrane interactions of the two C_2 domains of synaptotagmin. *J. Biol. Chem.* **269**, 31115–31123 (1994).
- Desai, R. C. et al. The C_2B domain of synaptotagmin is a Ca^{2+} -sensing module essential for exocytosis. *J. Cell Biol.* **150**, 1125–1136 (2000).
- Nonet, M. L., Grundahl, K., Meyer, B. J. & Rand, J. B. Synaptic function is impaired but not eliminated in *C. elegans* mutants lacking synaptotagmin. *Cell* **73**, 1291–1305 (1993).
- Broadie, K., Bellen, H. J., DiAntonio, A., Littleton, J. T. & Schwarz, T. L. Absence of synaptotagmin disrupts excitation-secretion coupling during synaptic transmission. *Proc. Natl Acad. Sci. USA* **91**, 10727–10731 (1994).
- DiAntonio, A. & Schwarz, T. L. The effect on synaptic physiology of synaptotagmin mutations in *Drosophila*. *Neuron* **12**, 909–920 (1994).
- Littleton, J. T., Stern, M., Perin, M. & Bellen, H. J. Calcium dependence of neurotransmitter release and rate of spontaneous vesicle fusions are altered in *Drosophila* synaptotagmin mutants. *Proc. Natl Acad. Sci. USA* **91**, 10888–10892 (1994).
- Geppert, M. et al. Synaptotagmin I: a major Ca^{2+} sensor for transmitter release at a central synapse. *Cell* **79**, 717–727 (1994).
- Jorgensen, E. M. et al. Defective recycling of synaptic vesicles in synaptotagmin mutants of *Caenorhabditis elegans*. *Nature* **378**, 196–199 (1995).
- Reist, N. E. et al. Morphologically docked synaptic vesicles are reduced in synaptotagmin mutants of *Drosophila*. *J. Neurosci.* **18**, 7662–7673 (1998).
- Littleton, J. T., Serano, T. L., Rubin, G. M., Ganetky, B. & Chapman, E. R. Synaptic function modulated by changes in the ratio of synaptotagmin I and IV. *Nature* **400**, 757–760 (1999).
- Wang, C. T. et al. Synaptotagmin modulation of fusion pore kinetics in regulated exocytosis of dense-core vesicles. *Science* **294**, 1111–1115 (2001).
- Ubach, J., Zhang, X., Shao, X., Sudhof, T. C. & Rizo, J. Ca^{2+} binding to synaptotagmin: how many Ca^{2+} ions bind to the tip of a C_2 -domain? *EMBO J.* **17**, 3921–3930 (1998).
- Li, C. et al. Ca^{2+} -dependent and -independent activities of neural and non-neural synaptotagmins. *Nature* **375**, 594–599 (1995).
- Zhang, X., Rizo, J. & Sudhof, T. C. Mechanism of phospholipid binding by the C_2A -domain of synaptotagmin I. *Biochemistry* **37**, 12395–12403 (1998).
- von Poser, C., Ichtchenko, K., Shao, X., Rizo, J. & Sudhof, T. C. The evolutionary pressure to inactivate. A subclass of synaptotagmins with an amino acid substitution that abolishes Ca^{2+} binding. *J. Biol. Chem.* **272**, 14314–14319 (1997).
- Loewen, C. A., Mackler, J. M. & Reist, N. E. *Drosophila* synaptotagmin I null mutants survive to early adulthood. *Genesis* **31**, 30–36 (2001).
- Rizo, J. & Sudhof, T. C. C_2 -domains, structure and function of a universal Ca^{2+} -binding domain. *J. Biol. Chem.* **273**, 15879–15882 (1998).
- Mackler, J. M., Drummond, J. A., Loewen, C. A., Robinson, I. M. & Reist, N. E. The C_2B Ca^{2+} -binding motif of synaptotagmin is required for synaptic transmission *in vivo*. *Nature* advance online publication, 7 July 2002 (doi:10.1038/nature00846).
- Neher, E. & Penner, R. Mice sans synaptotagmin. *Nature* **372**, 316–317 (1994).
- Voets, T. et al. Intracellular calcium dependence of large dense-core vesicle exocytosis in the absence of synaptotagmin I. *Proc. Natl Acad. Sci. USA* **98**, 11680–11685 (2001).
- Zhang, J. Z., Davletov, B. A., Sudhof, T. C. & Anderson, R. G. Synaptotagmin I is a high affinity receptor for clathrin AP-2: implications for membrane recycling. *Cell* **78**, 751–760 (1994).
- Hauke, V. & De Camilli, P. AP-2 recruitment to synaptotagmin stimulated by tyrosine-based endocytic motifs. *Science* **285**, 1268–1271 (1999).
- Burgoyne, R. D. & Morgan, A. Calcium sensors in regulated exocytosis. *Cell Calcium* **24**, 367–376 (1998).

30. DiAntonio, A., Parfitt, K. O. & Schwartz, T. L. Synaptic transmission persists in synaptotagmin mutants of *Drosophila*. *Cell* **73**, 1281–1290 (1993).
31. Martin, A. R. A further study of the statistical composition of the endplate. *J. Physiol.* **130**, 114–122 (1955).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank I. Inman for technical assistance; N. Reist, J. Li Bryan Stewart and B. Niemeyer for reagents and discussions; and N. Gay and the Department of Biochemistry (Cambridge) for resources. I.M.R. was supported by the American Heart Association, Western Affiliate, and is a Medical Research Council Career Development Award Fellow. The work was supported by the Muscular Dystrophy Association, by a Silvio Conti Center for Neuroscience Award from the National Institute of Mental Health, and by the National Institutes of Health (T.L.S.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.L.S. (e-mail: thomas.schwarz@tch.harvard.edu) or I.M.R. (e-mail: i.robinson@gen.cam.ac.uk).

The C₂B Ca²⁺-binding motif of synaptotagmin is required for synaptic transmission *in vivo*

J. M. Mackler*, J. A. Drummond†, C. A. Loewen*, I. M. Robinson† & N. E. Reist*

* Department of Anatomy and Neurobiology, Program in Molecular, Cellular, and Integrative Neuroscience, Colorado State University, Fort Collins, Colorado 80523, USA

† Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

Synaptotagmin is a synaptic vesicle protein that is postulated to be the Ca²⁺ sensor for fast, evoked neurotransmitter release¹. Deleting the gene for synaptotagmin (*sytn^{null}*) strongly suppresses synaptic transmission in every species examined², showing that synaptotagmin is central in the synaptic vesicle cycle. The cytoplasmic region of synaptotagmin contains two C₂ domains, C₂A and C₂B. Five, highly conserved, acidic residues in both the C₂A and C₂B domains of synaptotagmin coordinate the binding of Ca²⁺ ions^{3–5}, and biochemical studies have characterized several *in vitro* Ca²⁺-dependent interactions between synaptotagmin and other nerve terminal molecules⁶. But there has been no direct evidence that any of the Ca²⁺-binding sites within synaptotagmin are required *in vivo*. Here we show that mutating two of the Ca²⁺-binding aspartate residues in the C₂B domain (D_{416,418}N in *Drosophila*) decreased evoked transmitter release by >95%, and decreased the apparent Ca²⁺ affinity of evoked transmitter release. These studies show that the Ca²⁺-binding motif of the C₂B domain of synaptotagmin is essential for synaptic transmission.

The C₂B domain is critical for synaptotagmin function *in vivo*^{7–11}, yet most of the Ca²⁺-dependent, biochemical interactions of synaptotagmin were originally mapped to the C₂A domain¹². Thus, the importance of Ca²⁺ binding by the C₂B domain has been largely neglected. The recent finding that the C₂B domain mediates Ca²⁺-dependent interactions with phospholipids^{5,13} has heightened interest in Ca²⁺ binding by C₂B. However, the effects of mutations of the acidic residues that coordinate Ca²⁺ binding in the C₂B domain have not been examined at intact synapses. We mutated the C₂B aspartate residues to assess the importance of

the C₂B Ca²⁺-binding motif. Specifically, we mutated the third and fourth aspartate residues that coordinate Ca²⁺-binding in the C₂B domain of *Drosophila* synaptotagmin to asparagines (D_{416,418}N, referred to henceforth as B-D_{3,4}N, Fig. 1a) and expressed the mutant protein in a *sytn^{null}* line of *Drosophila*. Thus, the only source of synaptotagmin I in these lines (*P[sytn^{B-D3,4N}]*) was from the mutant transgene. As a negative control, we used the *sytn^{null}* line alone, and as a positive control, we used the same *sytn^{null}* line expressing synaptotagmin from a transgene bearing no mutations¹¹ (*P[sytn^{WT}]*).

Ca²⁺-evoked neurotransmitter release in *P[sytn^{B-D3,4N}]* mutant larvae was reduced to a level even less than that seen in the *sytn^{null}* negative control (Fig. 1b, c). Spontaneous and evoked potentials were recorded from third-instar muscle fibres as previously described¹⁴, except that evoked values were corrected for contamination by spontaneous release when appropriate¹⁵ (see Supplementary Information). When motor nerves of third-instar larvae were stimulated in standard HL3 saline¹⁶ containing 1.5 mM Ca²⁺, the mean amplitude of evoked excitatory junction potentials (EJPs) in muscles from *P[sytn^{WT}]* larvae was 25.0 ± 1.88 mV with no failures (*n* = 13), whereas mean EJP amplitude in *P[sytn^{B-D3,4N}]* larvae was only 0.21 ± 0.06 mV (*n* = 11) with a 78% failure rate. By comparison, mean EJP amplitude from *sytn^{null}* larvae was 1.70 ± 0.28 mV (*n* = 10) with a 15% failure rate.

In contrast to the decrease in evoked transmitter release, *P[sytn^{B-D3,4N}]* and *sytn^{null}* mutant synapses exhibited an increased rate of spontaneous vesicle fusion events (*P[sytn^{WT}]*, 2.9 ± 0.5 Hz; *P[sytn^{B-D3,4N}]*, 6.6 ± 1.4 Hz; *sytn^{null}*, 6.6 ± 0.8 Hz). The high frequency of miniature excitatory junction potentials (mEJPs) in mutant larvae shows that synaptic vesicles can fuse with the presynaptic membrane. The mean mEJP amplitudes for *P[sytn^{B-D3,4N}]* and *sytn^{null}* larvae were almost identical (0.99 ± 0.11 mV, *n* = 11 and 1.04 ± 0.08 mV, *n* = 10, respectively; *P* = 0.68), and slightly larger than that of *P[sytn^{WT}]* larvae (0.72 ± 0.09 mV, *n* = 11; *P* < 0.05). However, the frequency distribution curves of spontaneous mEJP amplitudes were remarkably similar among all genotypes (Fig. 1d), indicating that most vesicle fusions in all genotypes result in the same quantal sizes. The small number of larger events may represent either fusion of larger vesicles or an increased number of double and triple events owing to the increased mini frequency in the *sytn^{null}* and *P[sytn^{B-D3,4N}]* mutants. To further examine the fusibility of vesicles in *P[sytn^{B-D3,4N}]* mutant terminals, we focally applied a hypertonic solution (2.0 M sucrose) to elicit neurotransmitter release independent of nerve stimulation. Though variable, sucrose-stimulated release was comparable in *P[sytn^{WT}]* and *P[sytn^{B-D3,4N}]* larvae (Fig. 1e). Taken together, these data show that synaptic vesicles are capable of fusing with the presynaptic membrane. However, in *P[sytn^{B-D3,4N}]* mutants, nerve stimulation triggers the fusion of even fewer vesicles than in *sytn^{null}* mutants.

To determine the Ca²⁺ dependence of neurotransmitter release, we measured EJP amplitude at extracellular Ca²⁺ concentrations ranging from 0.4 to 5 mM for *P[sytn^{WT}]*, *P[sytn^{B-D3,4N}]* and *sytn^{null}* mutants (Fig. 1f). Although they differed greatly in maximal response, *P[sytn^{WT}]* and *sytn^{null}* fibres showed a similar Ca²⁺ dependence (Fig. 1f, g), with values of EC₅₀ (effector concentration for half-maximum response) of 1.05 and 0.93 mM, respectively. The maximal response in *P[sytn^{B-D3,4N}]* fibres was reduced even more than in *sytn^{null}* fibres. Moreover, *P[sytn^{B-D3,4N}]* fibres showed a significantly shifted Ca²⁺ dependence compared to *P[sytn^{WT}]* and *sytn^{null}*, with an EC₅₀ of 1.95 mM (Fig. 1f, g). To facilitate comparison of the *P[sytn^{WT}]* and *P[sytn^{B-D3,4N}]* mutants, the data were plotted as a percentage of the maximal response in each line (Fig. 1g). The rightward shift of the dose-response curve in the *P[sytn^{B-D3,4N}]* mutants shows that the apparent Ca²⁺ affinity of synaptic transmission is reduced by the B-D_{3,4}N mutation.

As evoked release was even lower in *P[sytn^{B-D3,4N}]* mutants than in *sytn^{null}* mutants, we examined whether the mutant protein dom-

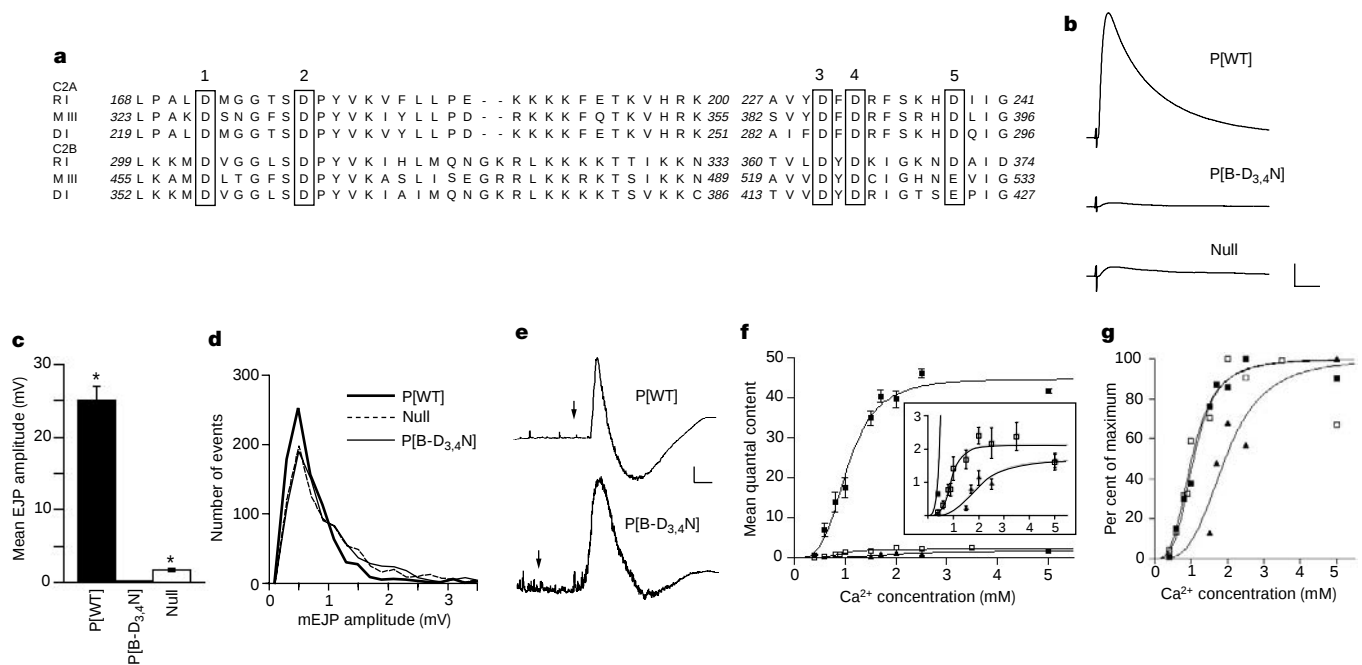


Figure 1 Fast, Ca^{2+} -evoked synaptic transmission is nearly abolished, and the apparent Ca^{2+} affinity of evoked transmitter release is decreased, in a C_2B Ca^{2+} -binding motif mutant. **a**, Partial amino-acid sequence of C_2A and C_2B domains of rat *syt I* (R I), mouse *syt III* (M III), and *Drosophila syt I* (D I). The five conserved acidic residues that coordinate the binding of multiple Ca^{2+} ions are boxed. **b**, Averages of 30 evoked excitatory junction potentials (EJPs) recorded in saline containing 1.5 mM Ca^{2+} . P[WT], *syt*^{null} expressing the P[*syt*^{WT}] transgene; P[B-D_{3,4}N], *syt*^{null} expressing the P[*syt*^{B-D_{3,4}N}] transgene; Null, *syt*^{null}. Scale bars: 5 mV, 20 ms. **c**, Mean EJP amplitude $\pm$ s.e.m. Asterisks indicate results that are significantly different from P[B-D_{3,4}N] ($P < 0.001$, Satterthwaite *t*-test).

d, Frequency distribution curves of miniature EJP (mEJP) amplitudes calculated from 850 individual events from each line compiled into 0.2-mV bins. **e**, Representative responses from P[WT] and P[B-D_{3,4}N] fibres to focal application (arrows) of 2.0 M sucrose. Scale bars: 5 mV, 1.25 s. **f**, Linear plot of $[\text{Ca}^{2+}]$ versus mean quantal content. Each point represents the mean quantal content of 6–10 muscle fibres from ≥ 3 larvae. Filled squares, P[*syt*^{WT}]; open squares, *syt*^{null}; triangles, P[*syt*^{B-D_{3,4}N}]. Error bars where visible indicate s.e.m. Inset: enlarged to show P[*syt*^{B-D_{3,4}N}] and *syt*^{null} curves. **g**, Ca^{2+} dose response data normalized to illustrate the rightward shift of the P[*syt*^{B-D_{3,4}N}] curve. Symbols as in **f**.

inantly inhibits synaptic transmission. We analysed evoked release in larvae that coexpressed native synaptotagmin and the transgenic B-D_{3,4}N mutant synaptotagmin. The control line only expressed synaptotagmin from the native gene, and not from the mutant transgene (see Supplementary Information). Under standard recording conditions, the control line (wild type, WT, Fig. 2a, b) displayed a mean EJP amplitude of $42.8 \text{ mV} \pm 1.1$ ($n = 9$). In contrast, evoked transmitter release in larvae expressing both the native and the B-D_{3,4}N mutant synaptotagmin (WT with B-D_{3,4}N,

Fig. 2a, b) was reduced by 53% to $20.3 \pm 3.2 \text{ mV}$ ($n = 11$, $P < 0.001$). Thus, the B-D_{3,4}N mutant protein inhibited synaptic transmission even in the presence of wild-type synaptotagmin.

We also mutated the first and second of the aspartate residues that coordinate Ca^{2+} -binding in the C_2B domain to asparagines (D_{356,362}N, referred to henceforth as B-D_{1,2}N, Fig. 1a). However, when crossed into the *syt*^{null} background, none of the embryos bearing the B-D_{1,2}N mutant transgene were able to hatch. The late embryonic lethality suggests that the B-D_{1,2}N mutation may be even

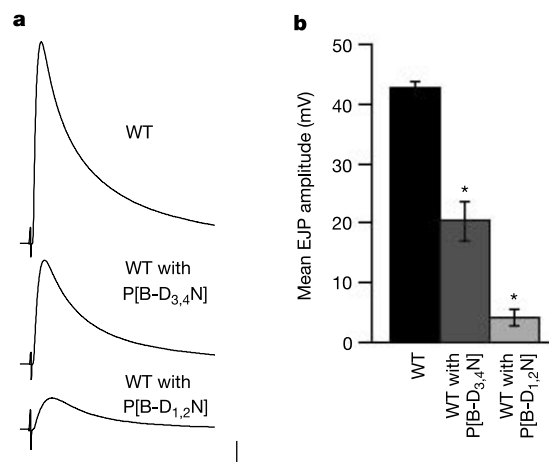


Figure 2 Expression of B-D_{3,4}N or B-D_{1,2}N mutant synaptotagmin inhibits synaptic transmission in otherwise wild-type synapses. **a**, Average of 30 evoked EJPs from three genotypes: WT, a line heterozygous for the native *syt* gene which contains the P[*syt*^{B-D_{3,4}N}] transgene but lacks transgene expression; WT with P[B-D_{3,4}N], a *syt*

heterozygote that expresses the B-D_{3,4}N mutant protein; WT with P[B-D_{1,2}N], a *syt* heterozygote that expresses the B-D_{1,2}N mutant protein. Scale bars: 5 mV, 20 ms. **b**, Mean EJP amplitude $\pm$ s.e.m. for all three lines. Results that are significantly different from WT are indicated by asterisks ($P < 0.001$, Satterthwaite *t*-test).

more detrimental to the organism than the B-D_{3,4}N mutation. Physiological analysis of larvae that coexpressed native synaptotagmin and the transgenic B-D_{1,2}N mutant synaptotagmin (WT with P[B-D_{1,2}N], Fig. 2a, b) support this hypothesis. The B-D_{1,2}N mutant protein inhibited evoked transmitter release by at least 90.5% (4.09 ± 1.33 mV, $n = 9$, $P < 0.0001$) in the presence of wild-type synaptotagmin. The only common feature between these mutations is that each is lacking two of the aspartate residues that coordinate Ca²⁺ binding in the C₂B domain of synaptotagmin. Thus, two independent mutations that disrupt the Ca²⁺-binding motif of C₂B each inhibit evoked transmitter release at *Drosophila* synapses.

To verify transgene expression in the *syt*^{null} background, we stained P[syt^{WT}] and P[syt^{B-D_{3,4}N}] larval whole mounts (Fig. 3a) and western blots of larval central nervous systems (CNS, Fig. 3b) with an anti-synaptotagmin antibody¹⁴. Both demonstrated abundant transgene expression. Thus in the *syt*^{null} background, the B-D_{3,4}N mutant protein is stably expressed in the larval nervous system and is appropriately targeted to neuromuscular junctions. To demonstrate transgene expression in the *syt* heterozygous background, similar western blots were analysed from the following lines: WT with P[syt^{B-D_{3,4}N}], WT with P[syt^{B-D_{1,2}N}], and their paired WT controls, each of which carry a transgene but do not express it (see Supplementary Information). Figure 3c shows that both of the transgene expressing lines (lanes 1 and 3) expressed

similar levels of synaptotagmin, whereas the control lines (lanes 2 and 4) expressed significantly less. Thus the mutant transgenes are also expressed in the *syt* heterozygous background. Coupled with the finding that evoked release was even lower in P[syt^{B-D_{3,4}N}] mutants than in *syt*^{null} mutants (which exhibit no detectable synaptotagmin expression¹⁴), the expression data suggest that the electrophysiological deficits observed in these C₂B Ca²⁺-binding motif mutants are unlikely to be a result of insufficient expression or mislocalization of the mutant transgenic protein.

The C₂B domain of synaptotagmin has been proposed to take part in vesicle recycling¹⁷. This hypothesis is supported by ultra-

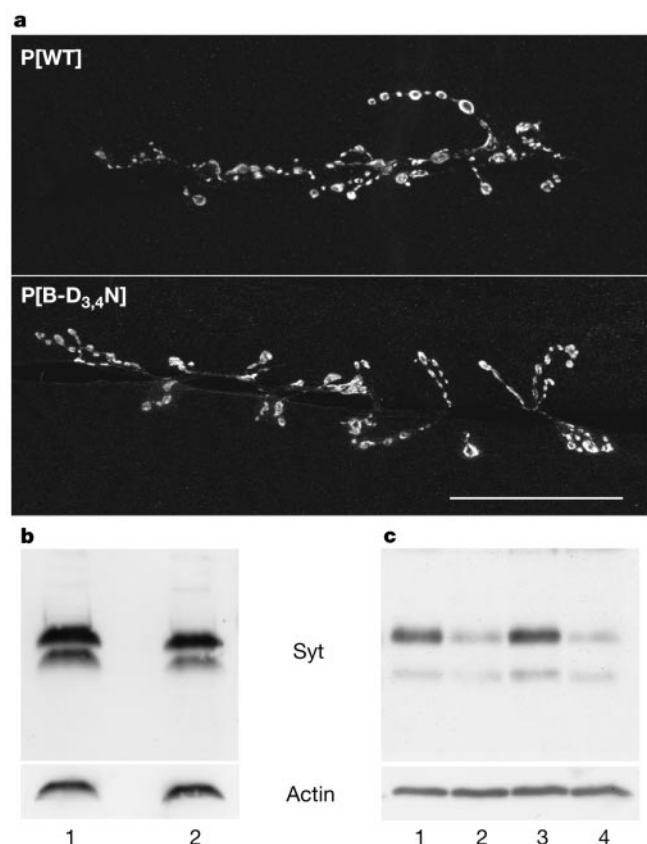


Figure 3 Synaptotagmin expression in the nervous system of P[syt^{WT}], P[syt^{B-D_{3,4}N}] and P[syt^{B-D_{1,2}N}] transgenic lines. **a**, Anti-synaptotagmin immunolabelling of P[syt^{WT}] and P[syt^{B-D_{3,4}N}] larvae shows that the wild-type and mutant transgenes are both abundantly expressed at neuromuscular junctions, the synapse used for all analyses. Scale bar, 50 μ m. **b**, Western blot showing synaptotagmin expression (seen as a doublet in *Drosophila*²⁸) in the CNS of *syt*^{null} larvae bearing synaptotagmin transgenes. Lane 1, P[syt^{WT}]; lane 2, P[syt^{B-D_{3,4}N}]. **c**, Western blot showing synaptotagmin transgene expression in *syt* heterozygotes. Lane 1, WT with P[B-D_{3,4}N]; lane 2, WT without P[B-D_{3,4}N] expression; lane 3, WT with P[B-D_{1,2}N]; lane 4, WT without P[B-D_{1,2}N] expression.

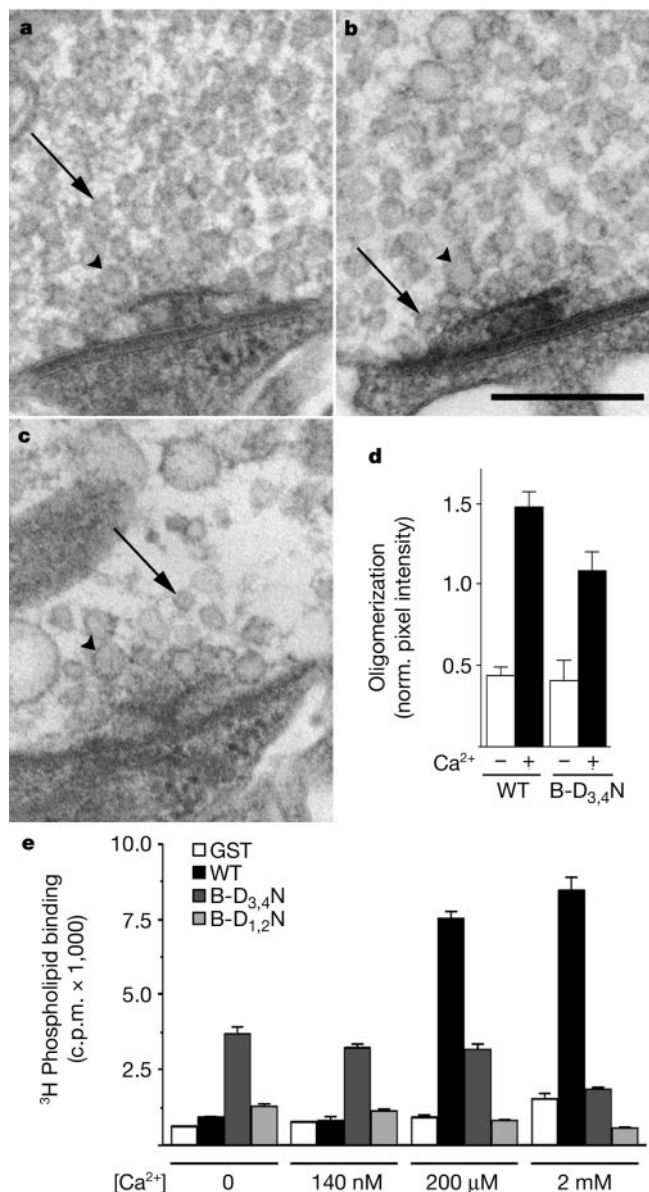


Figure 4 Whereas defects in recycling and Ca²⁺-dependent oligomerization cannot account for the severe decrease in synaptic transmission, decreased Ca²⁺-dependent phospholipid binding may account for this. Representative nerve terminals from neuromuscular junctions of P[syt^{WT}] (**a**), P[syt^{B-D_{3,4}N}] (**b**), and *syt*^{null} (**c**) third-instar larvae. Although there is an increase in the number of large vesicles (arrowheads), synaptic vesicles (arrows) are abundant in P[syt^{B-D_{3,4}N}] mutant nerve terminals. Scale bar, 250 nm. **d**, Ca²⁺-dependent oligomerization was decreased by ~25% in C₂A-C₂B domains bearing the B-D_{3,4}N mutation (mean $\pm$ s.e.m.). **e**, Ca²⁺-dependent phospholipid binding to GST-C₂B is severely reduced or abolished by the B-D_{3,4}N or B-D_{1,2}N mutation, respectively (GST, glutathione *S*-transferase). Background liposome binding is shown in the GST alone columns.

structural observations of *sytn^{null}* mutants^{18,19}; first-instar *sytn^{null}* mutants exhibit substantial depletion of synaptic vesicles and increased numbers of large vesicles, consistent with impaired recycling. If the electrophysiological deficits in our C₂B Ca²⁺-binding motif mutants were the result of a recycling defect, one would expect *P[sytn^{B-D3,4N}]* mutant terminals (Fig. 4b) to exhibit at least as much vesicle depletion as *sytn^{null}* mutant terminals (Fig. 4c). Qualitative observation of synaptic ultrastructure revealed that the *P[sytn^{B-D3,4N}]* mutants did exhibit some increase in large vesicles compared to controls (Fig. 4a, b); however, the abundance of synaptic vesicles in *P[sytn^{B-D3,4N}]* mutant terminals suggests that recycling is not the main defect in these mutants. Because the >95% decrease in evoked transmitter release in the *P[sytn^{B-D3,4N}]* mutant is not due to an absence of synaptic vesicles, the decreased release is probably a result of decreased release probability per vesicle.

In order to address the molecular mechanisms that may underlie the marked decrease in synaptic transmission caused by mutations in the C₂B Ca²⁺-binding motif, we tested wild-type and mutant *Drosophila* synaptotagmin I for Ca²⁺-dependent interactions *in vitro*. Syntaxin binding by the C₂B domain was not Ca²⁺ dependent²⁰, and the C₂B Ca²⁺-binding motif mutations did not alter this interaction (data not shown). As previously seen for rat synaptotagmin⁹, we found that the B-D_{3,4}N mutation caused a 25% reduction in Ca²⁺-dependent oligomerization (Fig. 4d, *P* = 0.052). Unlike the studies of the rat protein, however, the Ca²⁺-independent oligomerization was unchanged by the mutation. Ca²⁺-dependent phospholipid binding by the C₂B domain^{5,13}, on the other hand, was notably altered by the C₂B Ca²⁺-binding motif mutations (Fig. 4e). Compared to wild type, the B-D_{3,4}N mutation reduced Ca²⁺-dependent phospholipid binding by 65% at 200 μM Ca²⁺ and by 95% at 2 mM Ca²⁺ (*P* < 0.001), while increasing Ca²⁺-independent binding. The B-D_{1,2}N mutation reduced Ca²⁺-dependent phospholipid binding even more severely, down to levels similar to background binding to glutathione *S*-transferase (GST) alone (Fig. 4e). These mutations have also been shown to disrupt Ca²⁺-dependent interactions between rat *sytn I* and phospholipids²¹. Both the oligomerization assay and the phospholipid-binding assay show that the C₂B Ca²⁺-binding motif mutations disrupt Ca²⁺-dependent interactions of synaptotagmin.

The decreases in Ca²⁺-dependent phospholipid binding by the mutant C₂B domains closely reflect the severity of the electrophysiological deficits seen *in vivo*. The B-D_{3,4}N mutation severely decreased Ca²⁺-dependent phospholipid binding, and the B-D_{1,2}N mutation completely blocked it. This correlation suggests that Ca²⁺-dependent phospholipid binding by the C₂B domain may be a critical step in evoked transmitter release. Ca²⁺-dependent phospholipid binding by the C₂A domain is also thought to be important for evoked transmitter release. A mutation in the C₂A domain that decreased synaptotagmin's phospholipid binding *in vitro* also decreased evoked transmitter release²². However, alterations of the C₂A Ca²⁺-binding motif do not appear to inhibit evoked release *in vivo*²³. As the Ca²⁺-binding pockets of each C₂ domain are located in close proximity to one another^{4,24}, Ca²⁺ binding by the C₂B domain may facilitate membrane binding by the C₂A domain. Regardless of whether the C₂A and C₂B domains function cooperatively, our data—showing that mutation of two aspartates in C₂B inhibited evoked transmitter release by >95%—demonstrate that the C₂B Ca²⁺-binding motif is central to synaptotagmin function. Because synaptic vesicles in the *P[sytn^{B-D3,4N}]* mutant terminals were plentiful and fusion-competent, the dramatic decrease in evoked release is unlikely to result from a vesicle recycling or biogenesis defect. These results are (to our knowledge) the first to show a disruption of synaptic transmission caused by mutation of a Ca²⁺-binding motif, and demonstrate that the C₂B Ca²⁺-binding motif of synaptotagmin is essential for synaptic transmission *in vivo*. □

Methods

Site-directed mutagenesis and fly strains

Mutant *Drosophila sytn I* complementary DNA was generated by polymerase chain reaction (PCR) and used to transform *Drosophila* (see Supplementary Information). Either C₂B aspartate residues 416 and 418 (Fig. 1a, residues labelled 3 and 4) or 356 and 362 (Fig. 1a, residues labelled 1 and 2) were mutated to asparagines. *sytn^{AD4}* is a null mutation in the *Drosophila sytn I* gene²⁵. We used a homozygous *sytn^{AD4}* line, designated *sytn^{null}*, as the negative control. The transgenic lines used include: *P[sytn^{WT}]*; *P[sytn^{B-D3,4N}]*; WT with *P*[B-D_{3,4}N]; and WT with *P*[B-D_{1,2}N] (see Supplementary Information for genotypes).

Electrophysiology

Spontaneous and evoked potentials were recorded from third-instar muscle fibres as previously described¹⁴ in standard HL3 saline¹⁶ containing 1.5 mM Ca²⁺ unless otherwise indicated. For Ca²⁺ dose-response experiments, preparations were bathed in HL3 containing from 0.4 mM to 5 mM Ca²⁺ (without adjusting Mg²⁺). Transmitter release was also evoked by focal application of a hypertonic, 2.0 M sucrose solution to fibres in standard HL3 saline. To prevent overestimating nerve-evoked transmitter release in *sytn^{null}* and *P[sytn^{B-D3,4N}]* larvae, we restricted the measurement interval for EJP amplitudes and subtracted an estimate of the spontaneous release¹⁵ (see Supplementary Information). Failure rate was calculated as previously described²⁶.

Oligomerization

Soluble C₂A-C₂B domains were assayed for binding to immobilized GST-C₂A-C₂B domains according to the procedures of ref. 9 (see also Supplementary Information). Binding was wild type to wild type or mutant to mutant in buffer containing 2.5 mM CaCl₂ with or without 5 mM EGTA. Protein levels were quantified on Coomassie-blue-stained gels by comparison with bovine serum albumin (BSA) standards and were normalized to 12% of the total soluble protein used in the assay. There was no binding to GST alone. Binding assays were carried out three times.

Lipid binding

Lipid binding was assayed by the binding of liposomes to GST-C₂B fusion proteins²⁷. Recombinant protein bound to glutathione agarose beads was mixed with tritiated PS/PC liposomes (see Supplementary Information) in test solutions containing EGTA and CaCl₂ to give the desired free [Ca²⁺]. Liposome binding was quantified by scintillation counting. Data points were obtained in triplicate. Data shown are representative of three separate experiments.

Other techniques

Generation of recombinant proteins, western blot analysis, immunohistochemistry and electron microscopy experiments were carried out using standard techniques (see Supplementary Information).

Received 10 January; accepted 24 April 2002; doi:10.1038/nature00846.

Published online 7 July 2002.

- Augustine, G. J. How does calcium trigger neurotransmitter release? *Curr. Opin. Neurobiol.* **11**, 320–326 (2001).
- Lin, R. C. & Scheller, R. H. Mechanisms of synaptic vesicle exocytosis. *Annu. Rev. Cell Dev. Biol.* **16**, 19–49 (2000).
- Shao, X., Davletov, B. A., Sutton, R. B., Südhof, T. C. & Rizo, J. Bipartite Ca²⁺-binding motif in C₂ domains of synaptotagmin and protein kinase C. *Science* **273**, 248–251 (1996).
- Sutton, R. B., Ernst, J. A. & Brunger, A. T. Crystal structure of the cytosolic C₂A-C₂B domains of synaptotagmin III: Implications for Ca²⁺-independent SNARE complex interaction. *J. Cell Biol.* **147**, 589–598 (1999).
- Fernandez, I. et al. Three-dimensional structure of the synaptotagmin I C₂B-domain: Synaptotagmin I as a phospholipid binding machine. *Neuron* **32**, 1057–1069 (2001).
- Rizo, J. & Südhof, T. C₂-domains, structure and function of a universal Ca²⁺-binding domain. *J. Biol. Chem.* **273**, 15879–15882 (1998).
- Bommert, K. et al. Inhibition of neurotransmitter release by C₂-domain peptides implicates synaptotagmin in exocytosis. *Nature* **363**, 163–165 (1993).
- Fukuda, M. et al. Role of the C₂B domain of synaptotagmin in vesicular release and recycling as determined by specific antibody injection into the squid giant synapse preterminal. *Proc. Natl Acad. Sci. USA* **92**, 10708–10712 (1995).
- Desai, R. C. et al. The C₂B domain of synaptotagmin is a Ca²⁺-sensing module essential for exocytosis. *J. Cell Biol.* **150**, 1125–1136 (2000).
- Littleton, J. T. et al. *synaptotagmin* mutants reveal essential functions for the C₂B domain in Ca²⁺-triggered fusion and recycling of synaptic vesicles *in vivo*. *J. Neurosci.* **21**, 1421–1433 (2001).
- Mackler, J. M. & Reist, N. E. Mutations in the second C₂ domain of synaptotagmin disrupt synaptic transmission at *Drosophila* neuromuscular junctions. *J. Comp. Neurol.* **436**, 4–16 (2001).
- Li, C. et al. Ca²⁺-dependent and -independent activities of neural and non-neural synaptotagmins. *Nature* **375**, 594–599 (1995).
- Bai, J., Wang, P. & Chapman, E. R. C₂A activates a cryptic Ca²⁺-triggered membrane penetration activity within the C₂B domain of synaptotagmin I. *Proc. Natl Acad. Sci. USA* **99**, 1665–1670 (2002).
- Loewen, C. A., Mackler, J. M. & Reist, N. E. *Drosophila synaptotagmin I* null mutants survive to early adulthood. *Genesis* **31**, 30–36 (2001).
- Petersen, S. A., Fetter, R. D., Noordermeer, J. N., Goodman, C. S. & DiAntonio, A. Genetic analysis of glutamate receptors in *Drosophila* reveals a retrograde signal regulating presynaptic transmitter release. *Neuron* **19**, 1237–1248 (1997).
- Stewart, B. A., Atwood, H. L., Renger, J. J., Wang, J. & Wu, C.-F. Improved stability of *Drosophila* larval neuromuscular preparations in haemolymph-like physiological solutions. *J. Comp. Physiol. A* **175**, 179–191 (1994).
- Zhang, J. Z., Davletov, B. A., Südhof, T. C. & Anderson, R. G. Synaptotagmin I is a high affinity

- receptor for clathrin AP-2: implications for membrane recycling. *Cell* **78**, 751–760 (1994).
18. Jorgensen, E. M. *et al.* Defective recycling of synaptic vesicles in synaptotagmin mutants of *Caenorhabditis elegans*. *Nature* **378**, 196–199 (1995).
 19. Reist, N. E. *et al.* Morphologically docked synaptic vesicles are reduced in synaptotagmin mutants of *Drosophila*. *J. Neurosci.* **18**, 7662–7673 (1998).
 20. Kee, Y. & Scheller, R. H. Localization of synaptotagmin-binding domains on syntaxin. *J. Neurosci.* **16**, 1975–1981 (1996).
 21. Earles, C. A., Bai, J., Wang, P. & Chapman, E. R. The tandem C2 domains of synaptotagmin contain redundant Ca^{2+} binding sites that cooperate to engage t-SNAREs and trigger exocytosis. *J. Cell Biol.* **154**, 1117–1123 (2001).
 22. Fernandez-Chacon, R. *et al.* Synaptotagmin I functions as a calcium regulator of release probability. *Nature* **410**, 41–49 (2001).
 23. Robinson, I. M., Ranjan, R. & Schwarz, T. L. Synaptotagmins I and IV promote transmitter release independently of Ca^{2+} binding in the C2A domain. *Nature* advance online publication, 7 July 2002 (doi:10.1038/nature00915).
 24. Garcia, R. A., Forde, C. E. & Godwin, H. A. Calcium triggers an intramolecular association of the C2 domains in synaptotagmin. *Proc. Natl Acad. Sci. USA* **97**, 5883–5888 (2000).
 25. DiAntonio, A., Parfitt, K. D. & Schwarz, T. L. Synaptic transmission persists in synaptotagmin mutants of *Drosophila*. *Cell* **73**, 1281–1290 (1993).
 26. Del Castillo, J. & Katz, B. Quantal components of the end-plate potential. *J. Physiol.* **124**, 560–573 (1954).
 27. Davletov, B. A. & Sudhof, T. C. A single C2 domain from synaptotagmin I is sufficient for high affinity Ca^{2+} /phospholipid binding. *J. Biol. Chem.* **268**, 26386–26390 (1993).
 28. Littleton, J. T., Bellen, H. J. & Perin, M. S. Expression of synaptotagmin in *Drosophila* reveals transport and localization of synaptic vesicles to the synapse. *Development* **118**, 1077–1088 (1993).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank S. Royer, C. Williams and K. Mace for technical assistance, and R. Handa, K. Beam, J. Herbers, M. Tamkun and R. Aldrich for discussions about this manuscript. This work was supported by the National Science Foundation (N.E.R.), the Muscular Dystrophy Association (N.E.R.), the March of Dimes (N.E.R.) and an MRC Career Development Award (I.M.R.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.E.R. (e-mail: reist@lamar.colostate.edu).

Calorie restriction extends *Saccharomyces cerevisiae* lifespan by increasing respiration

Su-Ju Lin*, Matt Kaeberlein*†, Alex A. Andalis‡, Lori A. Sturtz§, Pierre-Antoine Defossez*†, Valeria C. Culotta§, Gerald R. Fink‡ & Leonard Guarente*

* Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

§ Department of Environmental Health Sciences, Johns Hopkins University School of Public Health, Baltimore, Maryland 21205, USA

Calorie restriction (CR) extends lifespan in a wide spectrum of organisms and is the only regimen known to lengthen the lifespan of mammals^{1–4}. We established a model of CR in budding yeast *Saccharomyces cerevisiae*. In this system, lifespan can be extended by limiting glucose or by reducing the activity of the glucose-sensing cyclic-AMP-dependent kinase (PKA)⁵. Lifespan extension in a mutant with reduced PKA activity requires Sir2 and NAD (nicotinamide adenine dinucleotide)⁵. In this study we explore how CR activates Sir2 to extend lifespan. Here we show

that the shunting of carbon metabolism toward the mitochondrial tricarboxylic acid cycle and the concomitant increase in respiration play a central part in this process. We discuss how this metabolic strategy may apply to CR in animals.

Ageing in yeast is regulated by *SIR2*. Deletion of *SIR2* shortens lifespan and overexpression of *SIR2* extends lifespan⁶. Sir2 exhibits an NAD-dependent histone deacetylase activity that is conserved among Sir2-family members and is required for chromatin silencing and lifespan extension^{7–9}. Recent studies show that *Caenorhabditis elegans* carrying extra copies of the *SIR2* orthologue, *sir-2.1*, also exhibit a longer lifespan¹⁰. The unusual NAD-requirement for the Sir2 deacetylase may link metabolic rate to silencing and lifespan¹¹.

Calorie restriction can be modelled in yeast by reducing the glucose content of the media from 2% to 0.5% (ref. 5). We first extended our earlier findings by testing whether Sir2 is required for the extension of replicative lifespan by 0.5% glucose. As shown in Fig. 1a, growth in 0.5% glucose extended lifespan (~25% increase) relative to the normal 2% glucose, and deletion of *SIR2* prevented this extension. We also examined whether Sir2-dependent ribosomal DNA silencing is increased. As shown in Fig. 1d, cells grown on 0.5% glucose medium exhibited enhanced silencing (using a *MET15* marker integrated at the ribosomal DNA¹² as an indicator). This enhanced silencing, like the extension of lifespan on 0.5% glucose, required Sir2 function.

How does CR increase Sir2 activity and extend lifespan? As shown in Fig. 2a, glucose is metabolized to pyruvate, at which point the pathway bifurcates into respiration and fermentation¹³. Respiration oxidizes the glucose to CO_2 generating 28 ATP molecules per molecule of glucose, whereas fermentation to ethanol generates only two ATP molecules per molecule of glucose¹⁴. When glucose levels are high, energy is in excess and fermentation is preferred. When glucose is limiting, respiration is preferred and carbon is shunted to the mitochondrial tricarboxylic acid (TCA) cycle, thereby increasing electron transport and respiration¹³.

To investigate whether this metabolic shift toward respiration occurs under our conditions of CR, we measured oxygen consumption rates of cells in 0.5% glucose. Figure 2b shows that respiration of cells grown in 0.5% glucose was significantly increased (~2-fold) compared to that of cells grown in 2% glucose. As a further test for this metabolic shift, we used a strain lacking *HXX2*, which encodes one of three hexokinases that introduce glucose into glycolysis. Deletion of *HXX2* is expected to mimic the effect of growth in low glucose and has been shown to extend lifespan⁵. Transcriptional profiling of the yeast genome indicates a highly significant overlap ($P = 3 \times 10^{-118}$) in the transcriptional changes caused by growth in 0.5% glucose and by deletion of *HXX2* (Fig. 3a). Similar to growth in 0.5% glucose, deletion of *HXX2* also increased the respiration rate significantly (~3-fold) (Fig. 2b).

Is the metabolic shift toward respiration required for CR-mediated lifespan extension? If so, elimination of electron transport ought to prevent the extension in lifespan. Thus, the gene encoding cytochrome c1, *CYT1*, was deleted, and lifespan analysis was performed (Fig. 2c). Calorie restriction failed to extend lifespan in the *cyt1Δ* mutant, suggesting that the metabolic shift toward respiration is necessary for lifespan extension mediated by CR.

We then investigated whether the metabolic shift toward respiration is sufficient for increased lifespan. Overexpression of the transcription factor Hap4 has been shown to cause a switch of metabolism from fermentation toward respiration¹⁵. Hap4 is expected to activate many genes involved in mitochondrial respiration^{16,17}. Transcription profiling indeed showed that many of the genes that were upregulated more than 2-fold by Hap4 overexpression fell into this category (Fig. 3c; Supplementary Information Table 2). Overexpression of Hap4 significantly extended the lifespan (~35% increase) of cells grown in 2% glucose (Fig. 1b). Similar to CR, this extension required Sir2 (Fig. 1b). Hap4 overexpression also

† Present addresses: Longevity Inc., Medford, Massachusetts 02155, USA (M.K.); and CNRS UMR 5665, Ecole Normale Supérieure de Lyon, France (P.-A.D.).

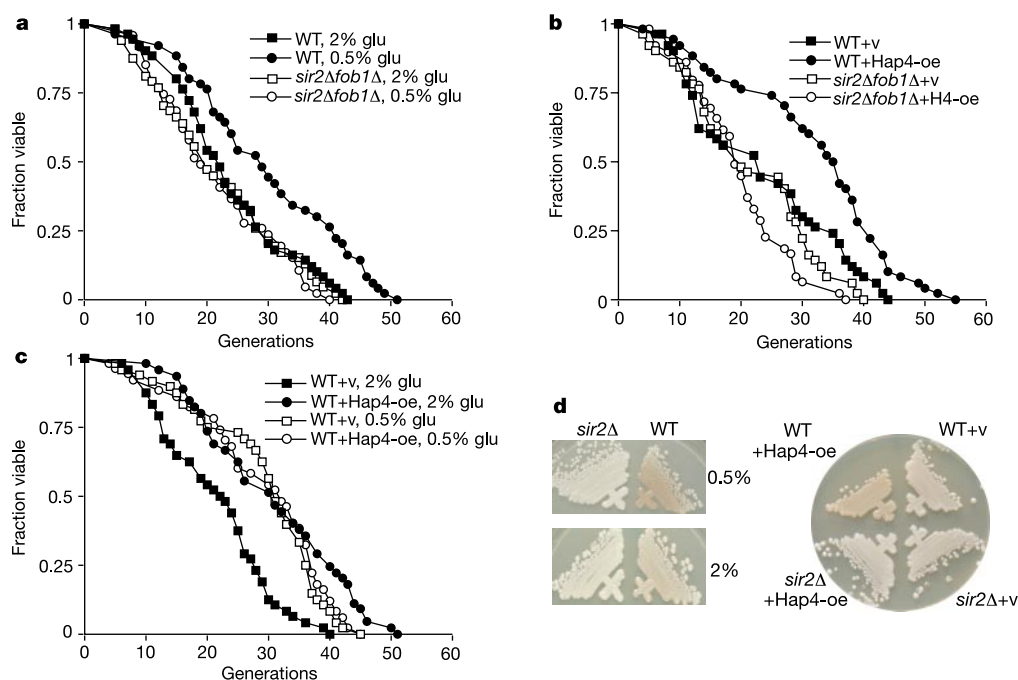


Figure 1 Calorie restriction (CR) and Hap4 overexpression extend lifespan and enhance rDNA silencing in a Sir2-dependent manner. **a**, Lifespan analysis of cells grown on media containing 0.5% and 2% glucose (glu). In 2% glucose, average lifespans in numbers of cell divisions: wild type (WT), 23.3; *sir2Δfob1Δ*, 21.5. In 0.5% glucose, average lifespans: WT, 29.4; *sir2Δfob1Δ*, 21.3. **b**, Lifespan analysis of wild-type and *sir2Δfob1Δ* mutant cells overexpressing Hap4. Average lifespans: WT + v, 23.2; WT + Hap4-oe, 31.9; *sir2Δfob1Δ* + v, 21.5; *sir2Δfob1Δ* + Hap4-oe, 19.6. The short lifespan of a *sir2ΔFOB1* strain was also unaffected by CR or by Hap4 overexpression (not shown). **c**, Lifespan analysis of cells overexpressing Hap4 on media containing 0.5% and 2% glucose. In 2% glucose, average lifespans: WT + v, 21.3; WT + Hap4-oe, 30.6. In 0.5%

glucose, average lifespans: WT + v, 29.1; WT + Hap4-oe, 29. For **a**, **b** and **c**: WT indicates PSY316 wild-type cells; +v indicates cells carrying a control vector; +Hap4-oe indicates cells overexpressing Hap4; *sir2Δfob1Δ* represents an isogenic derivative of PSY316. **d**, Silencing of a *MET15* marker at the rDNA locus. Accumulation of brown pigment indicates increased silencing. Left panels, cells were streaked onto rich medium containing 0.1% PbNO₃ and 0.5% (top) or 2% (bottom) glucose then grown for 5 d at 30 °C. Right panel, cells were streaked onto rich medium containing 0.1% PbNO₃ and 2% glucose then grown for 5 d at 30 °C. WT indicates JS237 wild-type cells; +v indicates cells carrying a control vector; +Hap4-oe indicates cells overexpressing Hap4; *sir2Δ* indicates JS218, an isogenic derivative of JS237.

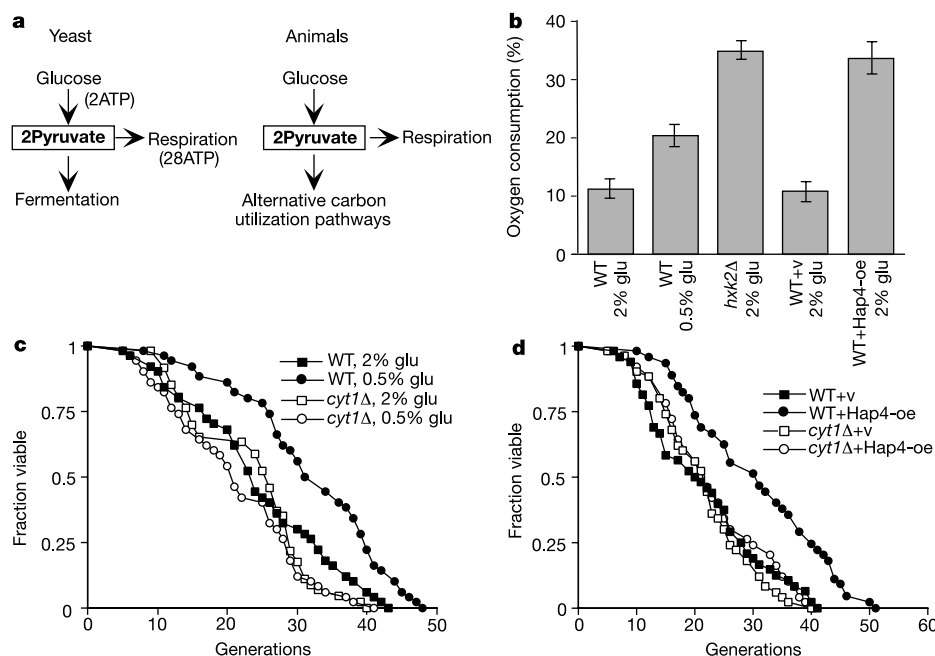


Figure 2 Respiration is required for calorie restriction (CR) and Hap4-overexpression-mediated lifespan extension. **a**, Schematic presentation of glucose utilization. During glycolysis, two molecules of pyruvate and ATP are generated per molecule of glucose, whereas respiration yields 28 ATP molecules per molecule of glucose. In mammals, pyruvate is dissimilated by respiration. In both systems, pyruvate is also used through alternative pathways such as synthesis of fatty acid and storage carbohydrates. **b**, Oxygen consumption measurement. Cells were grown in 2% or 0.5% glucose respectively as substrates. Results shown here represent the

average of 2 or 3 experimental trials, conducted in triplicate. Error bars show standard deviations. **c**, Deletion of *CYT1* prevents lifespan extension by CR. In 2% glucose, average lifespans: WT, 24.1; *cyt1Δ*, 23.3. In 0.5% glucose, average lifespans: WT, 31.9; *cyt1Δ*, 21. **d**, Deletion of *CYT1* prevents lifespan extension by Hap4 overexpression. Average lifespans: WT + v, 21; WT + Hap4-oe, 30.6; *cyt1Δ* + v, 21.5; *cyt1Δ* + Hap4-oe, 23. For **b**, **c** and **d**: WT indicates PSY316 wild-type cells; +v indicates cells carrying a control vector; +Hap4-oe indicates cells overexpressing Hap4; *cyt1Δ* and *hxx2Δ* represent isogenic derivatives of PSY316.

increased Sir2-dependent rDNA silencing (Fig. 1d). Thus, shifting metabolism toward respiration in cells grown in 2% glucose is sufficient to trigger the same effects on lifespan and silencing as does CR.

Three further findings reinforce the conclusion that the shift toward respiration induced by CR and by Hap4 overexpression is what extends lifespan. First, Hap4 overexpression did not synergize with growth on 0.5% glucose for the extension of lifespan (Fig. 1c). Second, Hap4 overexpression failed to extend lifespan in a respiration-deficient *cyt1Δ* mutant (Fig. 2d). Third, Hap4 overexpression resulted in a substantial increase in the respiration rate (~3-fold), similar to the increase observed for the *hxx2Δ* mutant (Fig. 2b).

We also compared the transcriptional effects of Hap4 overexpression to CR (the 124 genes co-regulated by 0.5% glucose and *hxx2Δ* mutant; Supplementary Information Table 1). As shown in Fig. 3b, of the 255 genes upregulated more than 2-fold by Hap4 overexpression (Supplementary Information Table 2), 55 genes were similarly regulated by CR (Supplementary Information Table 3). This overlap is statistically significant ($P = 1.5 \times 10^{-33}$). However, many of the transcriptional changes observed by overexpressing Hap4 were not observed during CR, and in fact, *HAP4* gene expression is not induced by CR. These observations suggest that transcriptional mechanisms unrelated to Hap4 or post-transcriptional mechanisms may also contribute to the activation of respiration by CR.

Other studies have suggested that CR functions to extend lifespan by increasing resistance to reactive oxygen species (ROS) and/or by reducing the production of ROS³. It thus seemed possible that the shift toward respiration under CR could extend lifespan by inducing resistance to ROS. However, expression of most antioxidant genes was not increased by Hap4 overexpression or by CR (Fig. 4a; Supplementary Information Table 4). Because a few genes, such as *SOD2* (mitochondrial superoxide dismutase), were slightly upregulated by Hap4 overexpression, we directly investigated whether CR or Hap4 overexpression increases resistance to oxidative stress. As shown in Fig. 4b, both CR and Hap4 overexpression, if anything, resulted in a slight decrease in resistance to paraquat. Further, while deletion of *SOD2* conferred a high degree of sensitivity to paraquat, deletion of *SIR2* had no effect. Also shown in Fig. 4c, CR and Hap4 overexpression slightly decreased resistance to hydrogen peroxide. As a control, deletion of *YAP1* gave rise to hypersensitivity to hydrogen peroxide, but, as for paraquat, deletion of *SIR2* did not. These findings all suggest that the Sir2-mediated extension of lifespan during CR is not due to an increase in resistance to oxidative stress.

How does this metabolic shift toward respiration during CR activate Sir2? First, the higher yield of ATP per input carbon allows a slower rate of glycolysis. Second, the increase in respiration yields a higher rate of electron transport, thereby increasing the re-oxidation of NADH to NAD in the mitochondria. This may increase the NAD/NADH ratio in the mitochondria, a change that can be transmitted to the cytosol by a shuttle that moves redox equivalents across the mitochondrial membrane¹⁸. It seems that the generation of NAD in the mitochondria is important, because disruption of the electron transport chain prevents the extension in lifespan by CR. Thus, we infer that the slowing of carbon flow through glycolysis, an increase in the NAD/NADH ratio, or both, contribute to the activation of Sir2 and the extension of lifespan.

Because CR in yeast does not increase resistance to oxidative stress, ROS are not limiting for the replicative lifespan of mother cells and are not central to the extension of lifespan by CR. ROS do, however, limit the long-term survival of stationary phase yeast cells^{19,20}, and may also play a role in the ageing of post-mitotic cells in *Drosophila melanogaster* and *C. elegans*^{21–23}. It is therefore intriguing that the replicative ageing of yeast mother cells and the post-mitotic ageing in *C. elegans* are both regulated by the levels of Sir2 proteins^{6,10}.

Our results are in contrast to the frequent suggestion that CR functions by slowing metabolism and thereby slowing the generation of ROS; in fact, data addressing whether CR actually changes the rate of respiration in animals is conflicting. In animals, respiration is essential for viability and mutations that reduce electron transport slow glycolysis and development and also lengthen lifespan²⁴. This is consistent with one of our proposed mechanisms for CR in yeast, that shifting metabolism toward respiration extends the lifespan by slowing glycolysis. In mammals, major pathways of carbon utilization, other than the TCA cycle, include the synthesis of fatty acids and storage carbohydrates¹³ (Fig. 2a). These alternative pathways (perhaps analogous to fermentation) are likely to be lowered to a greater degree during CR than the energy-producing TCA cycle.

We note that expression profiling of mouse skeletal muscle suggests that CR induces gluconeogenesis²⁵. This change is diagnostic of cells in a state of ATP excess, consistent with the hypothesis that CR diverts a higher fraction of the carbon into the TCA cycle. Finally, the increase in anti-oxidant enzymes that is reported to occur during CR in animals may be a result of an increase in respiration rather than a cause of the observed longevity. These considerations suggest that the essence of our models for CR in yeast could apply to animals. □

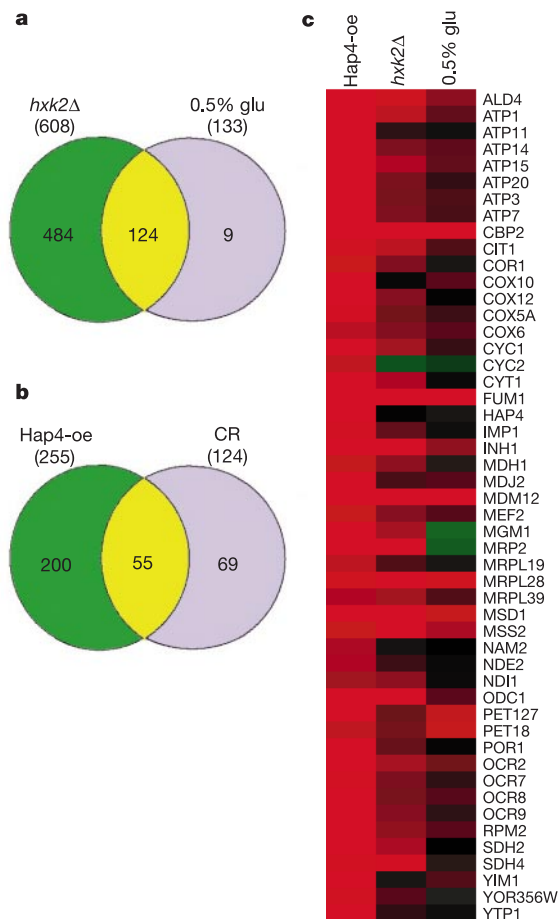


Figure 3 Gene expression profile analysis of cells under calorie restriction (CR) and overexpressing Hap4. **a**, Venn diagram of genes regulated by deletion of *HXX2* and growth in 0.5% glucose. Of the 133 genes differentially regulated by 0.5% glucose, 124 are similarly regulated by deletion of *HXX2*. **b**, Venn diagram of genes regulated by CR (0.5% glucose and *hxx2Δ*) and Hap4 overexpression (Hap4-oe). Of the 124 genes differentially regulated by CR, 55 genes are similarly regulated by overexpressing Hap4. **c**, Cluster analysis of genes regulated by CR (0.5% glucose and *hxx2Δ*) and Hap4 overexpression (Hap4-oe). Red represents genes that are upregulated and green represents genes that are downregulated.

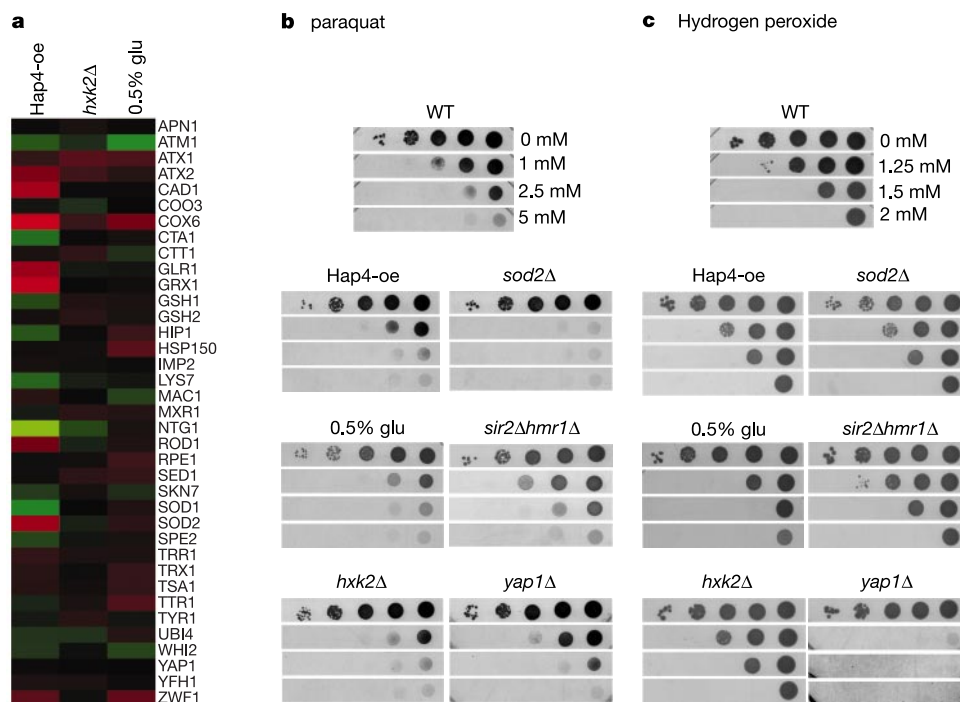


Figure 4 Calorie restriction (CR) and Hap4 overexpression do not increase the oxidative stress response. **a**, Cluster analysis of expression profile of genes involved in oxidative stress response in calorie-restricted (0.5% glucose and *hxx2Δ*) and Hap4-overexpressing (Hap4-oe) cells. Red represents genes that are upregulated and green represents genes that are downregulated. **b**, Paraquat sensitivity tests. **c**, Hydrogen peroxide sensitivity

tests. For **b** and **c**, serial 10-fold dilutions of cells were spotted onto synthetic complete media containing different concentrations of paraquat or hydrogen peroxide then grown at 30 °C for 2 d. The *HMR1* gene was deleted in the *sir2Δ* mutant to eliminate mating type effects. WT indicates PSY316 wild type; Hap4-oe indicates WT cells overexpressing Hap4; *hxx2Δ*, *sod2Δ*, *sir2Δhmr1Δ* and *yap1Δ* represent isogenic derivatives of PSY316.

Methods

Strains and plasmids

Strains PSY316 (*MATα*, *ura3-52*, *leu2-3*, 112, *his3-Δ200*, *ade2-101*, *lys2-801*, *RDN1::ADE2*), JS237 (*MATα*, *his3Δ200*, *met15Δ1*, *ura3-167*, *leu2Δ1*, *RDN1::Ty1-met15*) and JS218 (*MATα*, *his3Δ200*, *met15Δ1*, *ura3-167*, *leu2Δ1*, *RDN1::Ty1-met15*, *sir2Δ::HIS3*) have been previously described^{5,12}.

The integrating ppp81 (*LEU2*) vector, which allows gene overexpression under an *ADH1* promoter, was provided by P. Park. The Hap4 overexpression construct pADH1-HAP4 was made as follows: a pair of oligo-nucleotides were designed to amplify the *HAP4* coding region (from start codon to stop codon) by polymerase chain reaction (PCR) using Pfu DNA polymerase. This pair of oligo-nucleotides also added a *NotI* site to the 5' end and a *NheI* site to the 3' end of the *HAP4* gene. After PCR amplification, DNA was digested with *NotI* and *NheI*, then ligated to ppp81 digested with the same enzymes resulting in pADH1-HAP4. DNA sequencing analysis of pADH1-HAP4 showed that no mutation was introduced into the *HAP4* gene during PCR amplification. This construct was further verified by its ability to suppress the growth defects of a *hap4Δ* mutant on non-fermentable carbon.

WT + Hap4-oe strains were made by integrating *XcmI* linearized pADH1-HAP4 into the *LEU2* locus. WT + v strains were made by integrating *XcmI* linearized ppp81 into the *LEU2* locus. All gene deletions in this study were constructed by replacing the wild-type genes with the *Kan^r* marker as described²⁶, and verified by PCR using oligo-nucleotides flanking the genes of interest.

Lifespan, silencing and paraquat sensitivity assays

Lifespan analyses were carried out as previously described⁵. All lifespan analyses in this study were carried out on YPD plates at least twice independently with more than 45 cells per strain per experiment. Results from a single experiment are shown.

Silencing assays were carried out by streaking cells onto YPD media containing 0.1% PbNO₃ and 2% or 0.5% glucose, then grown for 5 d at 30 °C. Serial dilution assays were performed by growing cells overnight in YPD at 30 °C. Cells were then diluted 50-fold in fresh YPD and cultured for 4–5 h. For each strain, a series of 10-fold dilutions were prepared in fresh synthetic media over a range of concentrations from 10⁻¹ to 10⁻⁵, relative to the initial culture. 5 μl of the original culture and each dilution were spotted sequentially onto the appropriate media. The cells were then grown at 30 °C for 2 d before visualization. Paraquat media were prepared by supplementing synthetic complete media with 0, 1, 2.5, or 5 mM paraquat (Methyl Viologen, Sigma). Hydrogen peroxide media were prepared by supplementing synthetic complete media with 0, 1.25, 1.5, 2 mM hydrogen peroxide (H₂O₂, Sigma).

Oxygen consumption assay

For oxygen consumption experiments, cells were first grown overnight in YPD media

containing either 2% or 0.5% glucose, then diluted into fresh media the next day. For cells grown in 2% glucose, cultures were diluted to an optical density at 600 nm (OD₆₀₀) of 0.2. Cells grown in 0.5% glucose were diluted to OD₆₀₀ = 0.05. Growth of cultures was continued at 30 °C with shaking for approximately 6–8 h until the final OD₆₀₀ of cultures grown in 2% glucose was 0.8 and that of cultures grown in 0.5% glucose was 0.2. Cells were then harvested, washed and resuspended in 500 μl of fresh YPD media containing either 2% or 0.5% glucose. Oxygen consumption was measured using a YSI Model 5300 Biological Oxygen Monitor equipped with a Clark-type oxygen electrode (Yellow Springs Instrument) and readings were recorded every 15 s. For WT + v and WT + Hap4-oe strains, 2.5 × 10⁸ cells were assayed. For WT cells grown in 2% and 0.5% glucose and the *hxx2Δ* strain, 1.0 × 10⁹ cells were used. Results are reported as per cent O₂ consumed per minute per 1 × 10⁹ cells.

Microarray analysis

Slides for microarray analysis were prepared by PCR amplification of yeast open reading frames (ORFs) using the Research Genetics Yeast Gene Pairs primer set. All ORFs were spotted onto glass slides coated with poly-L-lysine using a Genemachines Omnigrid Microarrayer and post-processed²⁷.

To obtain messenger RNA for microarray analysis, cells were grown in YPD overnight and diluted 100-fold into fresh YPD and incubated at 30 °C to an OD₆₀₀ between 0.6 and 0.8. For cells grown in low glucose, cultures were incubated overnight in YPD containing 0.5% glucose then diluted 250-fold and grown to OD₆₀₀ = 0.2. Cells were then washed in water and frozen in dry ice/ethanol before storage at -80 °C.

Total RNA was obtained by the acid-phenol method. mRNA was purified from total RNA using the PolyATtract mRNA isolation system (Promega). 2 μg of mRNA was labelled for each sample using the reverse transcription and amino-allyl coupling method described at <http://www.microarrays.org>. Reference samples of wild-type PSY316 mRNA harvested from cells grown in YPD at 30 °C were labelled with Cy3 and experimental samples were labelled with Cy5. Labelled samples were combined and hybridized overnight at 42 °C. After hybridization, slides were washed²⁷ and scanned using a Packard BioScience 3000 scanner. Images were processed using the MolecularWare DigitalGENOME (Cambridge, MA) software package. All microarray data sets are available as Microsoft Excel files at <http://web.mit.edu/biology/guarente/arrays/kaeberlein>.

A scaling factor was calculated and applied in each microarray experiment so that the median ratio value of all experiments was equal to 1. Genes were defined as differentially regulated if the Cy5/Cy3 ratio was greater than 1.5 or less than 0.667 (-1.5-fold) in both of two independent experiments. Reported values are the mean values of duplicate experiments. For the *HXX2* deletion, a single microarray experiment was performed. Cluster analysis was performed using Cluster and visualized with TreeView²⁸. Red represents genes that are upregulated and green represents genes that are downregulated. Statistical significance of the overlap between regulated genes in different experiments

shown in Fig. 3 was calculated using a hypergeometric distribution. *P*-values were obtained from the online hypergeometric distribution calculator at <http://www.alewand.de/stattab/tabdiske.htm>.

Received 8 March; accepted 12 April 2002; doi:10.1038/nature00829.

- Weindruch, W. & Walford, R. L. *The Retardation Of Aging And Diseases By Dietary Restriction* (Thomas, Springfield, Illinois, 1998).
- Roth, G. S., Ingram, D. K. & Lane, M. A. Calorie restriction in primates: will it work and how will we know? *J. Am. Geriatr. Soc.* **47**, 896–903 (1999).
- Sohal, R. S. & Weindruch, R. Oxidative stress, caloric restriction, and aging. *Science* **273**, 59–63 (1996).
- Yu, B. P. *Modulation of Aging Processes by Dietary Restriction* (CRC Press, Boca Raton, Florida, 1994).
- Lin, S. J., Defossez, P. A. & Guarente, L. Requirement of NAD and *SIR2* for life-span extension by calorie restriction in *Saccharomyces cerevisiae*. *Science* **289**, 2126–2128 (2000).
- Kaeberlein, M., McVey, M. & Guarente, L. The *SIR2/3/4* complex and *SIR2* alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.* **13**, 2570–2580 (1999).
- Smith, J. S. *et al.* A phylogenetically conserved NAD⁺-dependent protein deacetylase activity in the Sir2 protein family. *Proc. Natl Acad. Sci. USA* **97**, 6658–6663 (2000).
- Landry, J. *et al.* The silencing protein *SIR2* and its homologs are NAD-dependent protein deacetylases. *Proc. Natl Acad. Sci. USA* **97**, 5807–5811 (2000).
- Imai, S., Armstrong, C. M., Kaeberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
- Tissenbaum, H. A. & Guarente, L. Increased dosage of a *sir-2* gene extends lifespan in *Caenorhabditis elegans*. *Nature* **410**, 227–230 (2001).
- Guarente, L. Sir2 links chromatin silencing, metabolism, and aging. *Genes Dev.* **14**, 1021–1026 (2000).
- Smith, J. S. & Boeke, J. D. An unusual form of transcriptional silencing in yeast ribosomal DNA. *Genes Dev.* **11**, 241–254 (1997).
- Pronk, J. T., Yde Steensma, H. & Van Dijken, J. P. Pyruvate metabolism in *Saccharomyces cerevisiae*. *Yeast* **12**, 1607–1633 (1996).
- Stryer, L. *Biochemistry* (Freeman, New York, 1995).
- Blom, J., De Mattos, M. J. & Grivell, L. A. Redirection of the respiro-fermentative flux distribution in *Saccharomyces cerevisiae* by overexpression of the transcription factor Hap4. *Appl. Environ. Microbiol.* **66**, 1970–1973 (2000).
- de Winder, J. H. & Grivell, L. A. Global regulation of mitochondrial biogenesis in *Saccharomyces cerevisiae*. *Prog. Nucleic Acid Res. Mol. Biol.* **46**, 51–91 (1993).
- Forsburg, S. L. & Guarente, L. Identification and characterization of *HAP4*: a third component of the CCAAT-bound *HAP2/HAP3* heteromer. *Genes Dev.* **3**, 1166–1178 (1989).
- Bakker, B. M. *et al.* Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* **25**, 15–37 (2001).
- Longo, V. D., Gralla, E. B. & Valentine, J. S. Superoxide dismutase activity is essential for stationary phase survival in *Saccharomyces cerevisiae*. Mitochondrial production of toxic oxygen species *in vivo*. *J. Biol. Chem.* **271**, 12275–12280 (1996).
- Fabrizio, P., Pozza, F., Pletcher, S. D., Gendron, C. M. & Longo, V. D. Regulation of longevity and stress resistance by Sch9 in yeast. *Science* **292**, 288–290 (2001).
- Melov, S. *et al.* Extension of life-span with superoxide dismutase/catalase mimetics. *Science* **289**, 1567–1569 (2000).
- Orr, W. C. *et al.* Extension of life-span by overexpression of superoxide dismutase and catalase in *Drosophila melanogaster*. *Science* **263**, 1128–1130 (1994).
- Taub, J. *et al.* A cytosolic catalase is needed to extend adult lifespan in *C. elegans daf-1* mutants. *Nature* **399**, 162–166 (1999).
- Feng, J., Bussiere, F. & Hekimi, S. Mitochondrial electron transport is a key determinant of lifespan in *Caenorhabditis elegans*. *Dev. Cell* **1**, 633–644 (2001).
- Lee, C. K., Klopp, R. G., Weindruch, R. & Prolla, T. A. Gene expression profile of aging and its retardation by caloric restriction. *Science* **285**, 1390–1393 (1999).
- Guldener, U., Heck, S., Fielder, T., Beinhausen, J. & Hegemann, J. H. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* **24**, 2519–2524 (1996).
- Hegde, P. *et al.* A concise guide to cDNA microarray analysis. *Biotechniques* **29**, 548–550, 552–554, 556 (2000).
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank D. McNabb and members of the Guarente laboratory for suggestions; J. Smith for providing strains; T. Galitski for his contributions in the development of microarrays and analytical software tools; and T. Ideker for suggestions with microarray analysis. This work was supported by grants to L.G. from the National Institute of Health (NIH), The Ellison Medical Foundation, The Seaver Institute, and the Howard and Linda Stern Fund. S.-J.L. is supported by a NRSA individual award. G.R.F. is supported by the NIH and is an American Cancer Society Professor of Genetics. A.A.A. is supported by the NIH Training Grant in Genomic Sciences, sponsored by the Biotechnology Process Engineering Center. V.C.C. and L.A.S. are supported by grants from the NIH.

Competing interests statement

The authors declare competing financial interests: details accompany the paper on Nature's website (<http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to L.G. (e-mail: leng@mit.edu).

Adenovirus oncoproteins inactivate the Mre11–Rad50–NBS1 DNA repair complex

Travis H. Stracker^{*†}, Christian T. Carson^{*†} & Matthew D. Weitzman^{*}

^{*} Laboratory of Genetics, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

[†] Graduate Program, Department of Biology, University of California, San Diego, California 92093, USA

In mammalian cells, a conserved multiprotein complex of Mre11, Rad50 and NBS1 (also known as nibrin and p95) is important for double-strand break repair, meiotic recombination and telomere maintenance^{1–4}. This complex forms nuclear foci and may be a sensor of double-strand breaks. In the absence of the early region E4, the double-stranded DNA genome of adenovirus is joined into concatemers too large to be packaged^{5,6}. We have investigated the cellular proteins involved in this concatemer formation

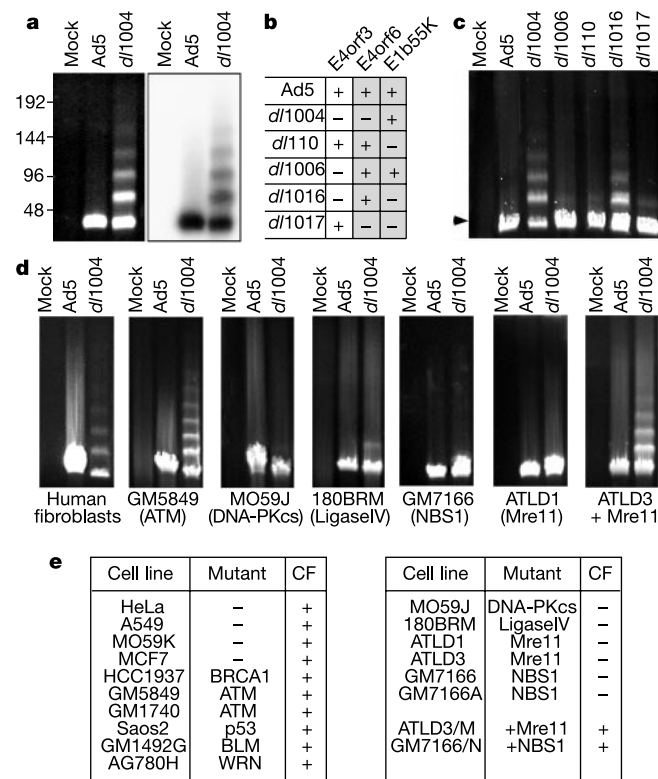


Figure 1 Adenovirus genome concatemer formation with wild-type (Ad5) and mutant viruses in human cell lines. **a**, Concatemers of Ad genomes. HeLa cells were either uninfected (mock) or infected with wild-type Ad5 or an E4-deleted virus (d1004) at an MOI of 25 p.f.u. per cell. DNA was prepared at 30 h.p.i. and analysed by PFGE. Left, gel stained with ethidium bromide with the positions of size markers indicated in kilobases. Right, Southern blot hybridization with an Ad probe. **b**, Genotype of mutant adenoviruses used for infections. Shading indicates that the E4orf3 and E1b55K proteins function as a complex. **c** Concatemer formation is observed when both E4orf3 and the E4orf3–E1b55K complex are inactivated. HeLa cells were infected with mutant viruses and analysed by PFGE. The position of the single linear viral genome is indicated. **d**, Cellular requirements for concatemer formation. Wild-type and mutant cell lines were infected with viruses and DNA was analysed by PFGE. Mutated genes are indicated in parentheses. **e**, Summary of concatemer formation (CF) in mutant cell lines. Reconstituted cell lines are shown at the bottom.

and how they are inactivated by E4 products during a wild-type infection. Here we show that concatemerization requires functional Mre11 and NBS1, and that these proteins are found at foci adjacent to viral replication centres. Infection with wild-type virus results in both reorganization and degradation of members of the Mre11–Rad50–NBS1 complex. These activities are mediated by three viral oncoproteins that prevent concatemerization. This targeting of cellular proteins involved in genomic stability suggests a mechanism for ‘hit-and-run’ transformation observed for these viral oncoproteins⁷.

We used pulsed-field gel electrophoresis (PFGE) to analyse viral DNA from cells infected with wild-type or mutants of adenovirus (Ad) type 5 (Fig. 1). For an E4-deleted virus (*dl1004*), a ladder of bands was observed that was shown by Southern hybridization to comprise Ad genome concatemers. Two of the E4 open reading frames (E4orf3 and E4orf6) encode redundant gene products, and either is sufficient for replication^{8–10} and for preventing concatemer formation⁵. E4orf6 forms a multifunctional complex with the viral E1b55K protein¹¹. Concatemers were also observed for a virus mutated in both E4orf3 and E1b55K (*dl1016*). This suggests that E4orf3 and the E4orf6–E1b55K complex can each prevent joining of viral genomes and that concatemers are formed only when both are inactivated.

We used human cell lines with mutations in DNA repair genes to examine cellular proteins required for concatemerization (Fig. 1d, e). Concatemers formed in cells with mutant p53, Bloom’s syndrome protein (BLM), Werner’s syndrome protein (WRN), breast cancer susceptibility protein 1 (BRCA1), and ataxia-telangiectasia mutated protein (ATM). In contrast, no concatemers were observed

in cells with mutant DNA ligase IV (ref. 12) and DNA-dependent protein kinase catalytic subunit (DNA-PKcs) as previously reported (ref. 6). Concatemerization was also absent in cells that lacked functional Mre11 and NBS1 proteins, but was restored in mutant cells reconstituted with wild-type genes. The Mre11–Rad50–NBS1 complex is therefore essential for concatemer formation and may represent a target of the E4 proteins.

We assessed the effect of Ad infection on the localization of Mre11, Rad50 and NBS1 by immunofluorescence (Fig. 2a). In uninfected cells, these proteins appear throughout the nucleoplasm but are also concentrated at sites of replication and at discrete nuclear structures corresponding to promyelocytic leukemia (PML) oncogenic domains (PODs), also known as nuclear domain 10 (ND10) (refs 13–15). Viral replication centres were visualized with antibodies to the viral DNA-binding protein (DBP). Early after infection the Mre11, Rad50 and NBS1 proteins were detected in nuclear speckles (Fig. 2b). At later times, decreased Mre11 staining suggested that this protein was downregulated by the virus, and immunoblotting confirmed a decreased abundance of the Mre11, Rad50 and NBS1 proteins during virus infection (Fig. 2c). This was caused by enhanced protein turnover, as shown by blocking protein synthesis with cycloheximide (Fig. 2d). Adding the proteasome inhibitor MG132 stabilized the amounts of Mre11 and Rad50 in infected cells (Fig. 2e). These results show that infection with wild-type Ad results in altered localization and proteasome-mediated degradation of members of the Mre11–Rad50–NBS1 complex.

In response to cellular double-strand breaks, Mre11, Rad50 and NBS1 form foci at sites of DNA damage^{4,16}. We examined Mre11–Rad50–NBS1 complexes after infection with an E4-deleted virus

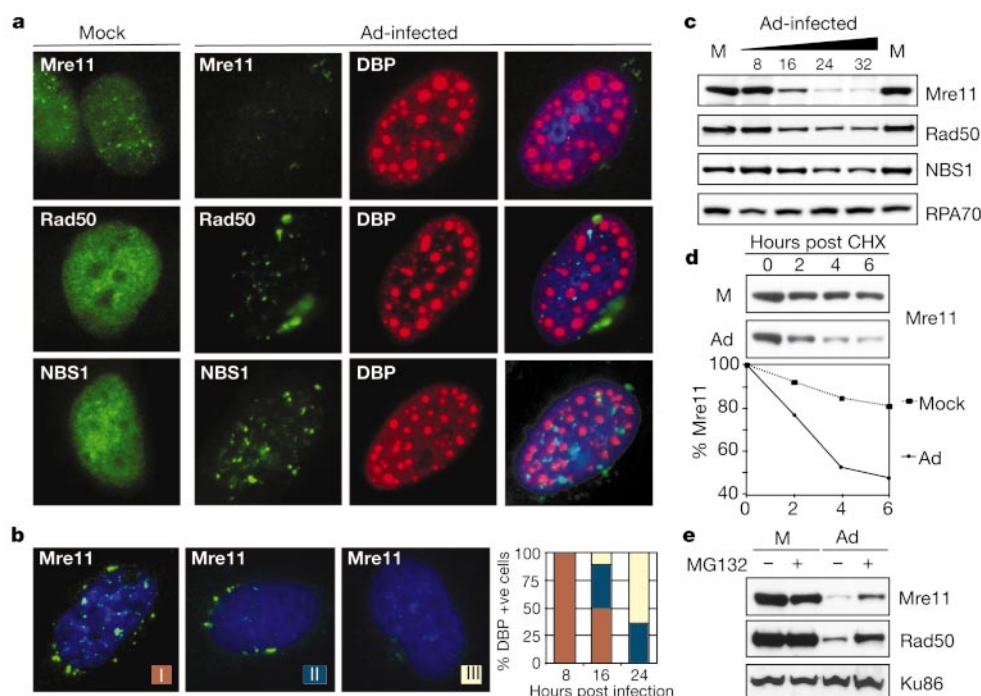


Figure 2 The Mre11, Rad50 and NBS1 proteins are altered during infection with Ad. **a**, Localization of the Mre11–Rad50–NBS1 complex during Ad infection. HeLa cells were either untreated (mock) or infected with Ad5 (MOI 25 p.f.u. per cell) and proteins were visualized at 18 h.p.i. by indirect immunofluorescence. Cells were stained with mouse monoclonal antibodies to Mre11 and Rad50, and rabbit polyclonal antibody to NBS1. Viral replication centres were visualized by staining for DBP. Merged images with DAPI staining of nuclei are shown on the right. **b**, Quantification of staining patterns. Three patterns of Mre11 staining were observed during viral infection and these were quantified over a time course. **c**, The Mre11 and Rad50 proteins are downregulated on infection with Ad. Lysates were prepared at the indicated times from HeLa cells that were either mock-

treated (M) or infected with Ad, and proteins were detected by immunoblotting. RPA70 was included as a loading control. **d**, Cycloheximide treatment (CHX) of HeLa cells shows active degradation of Mre11 during an Ad5 infection. Cells were incubated with 25 $\mu\text{g ml}^{-1}$ cycloheximide after 10 h of Ad5 infection or mock treatment (M), and lysates were prepared for immunoblotting at the indicated times after treatment. Quantification of Mre11 is shown below. **e**, Degradation of Mre11 and Rad50 is mediated by the proteasome. The proteasome inhibitor MG132 was added to mock-treated (M) or Ad-infected cells at 10 h.p.i. and lysates were prepared for immunoblotting at 24 h.p.i. Mre11 and Rad50 were stabilized by the proteasome inhibitor, whereas amounts of Ku86 remained unaltered.

(*dl1004*). All three components of the complex were found in foci surrounding viral replication centres (Fig. 3a). Colocalization was confirmed by deconvolution microscopy (Fig. 3b). Pulse-labelling showed that bromodeoxyuridine (BrdU) partially overlapped with NBS1; thus, the Mre11–Rad50–NBS1 complex was recruited to sites of active viral DNA replication. Immunoblotting showed that, in contrast to infection with the wild-type virus, the amounts of Mre11, NBS1 and Rad50 proteins were unaffected (Fig. 3c). A shift was observed in the migration of the NBS1 protein because of altered phosphorylation, as shown by phosphatase treatment and immunoprecipitation with a phospho-specific antibody (Fig. 3d). A mobility shift was observed in cells lacking DNA-PKcs (T.H.S. and M.D.W., unpublished data) and ATM (Fig. 3d), which indicates that additional kinases such as the Rad3 related protein (ATR) may be activated by infection with *dl1004*. NBS1 is phosphorylated in response to DNA-damaging agents^{17–19}, which suggests that virus infection in the absence of E4 triggers a cellular response to DNA damage.

We used viruses with mutations in specific E4 genes to define the viral proteins involved. Cells infected with a virus lacking E4orf6–E1b55K (*dl1017*) showed nuclear speckles of Mre11 and Rad50 but no degradation (Fig. 3e). Cells infected with a virus lacking E4orf1 to E4orf3 (Δ E4orf1–3) (*dl1006*) showed rapid degradation of the Mre11, Rad50 and NBS1 proteins but no nuclear speckles. Infec-

tions with the double mutant virus (*dl1016*) resulted in recruitment of the Mre11–Rad50–NBS1 complex to viral replication centres, coincident with formation of concatemers (Fig. 1c). Observations from immunofluorescence were supported by immunoblotting (Fig. 3f). These results link expression of E4orf3 to redistribution of the Mre11, Rad50 and NBS1 proteins and implicate E4orf6–E1b55K in degradation.

To investigate further the mechanisms of preventing concatemer formation, we transfected expression vectors for individual E4 gene products. In 293 cells, which contain integrated adenoviral E1 genes, Mre11 and Rad50 were degraded by E4orf6 (Fig. 4a) but were unaffected by an E4orf6 mutant (E4orf6.AXA) that does not form a functional complex with E1b55K²⁰. The E4orf6–E1b55K complex thus induces degradation of cellular Mre11, Rad50 and NBS1 proteins. Degradation may be mediated by a direct interaction and recruitment of ubiquitin ligase activity²¹, or alternatively the viral proteins might affect a cellular regulator of Mre11 complex stability. The E4orf3 protein is sufficient to disrupt POD–ND10 structures²² and to reorganize PML into nuclear tracks (Fig. 4b). On expression of E4orf3, Mre11, Rad50 and NBS1 were observed in the cytoplasm and in a speckled nuclear pattern that overlapped partially with PML tracks.

To determine whether the observed effects of the viral proteins on the Mre11–Rad50–NBS1 complex are necessary for preventing

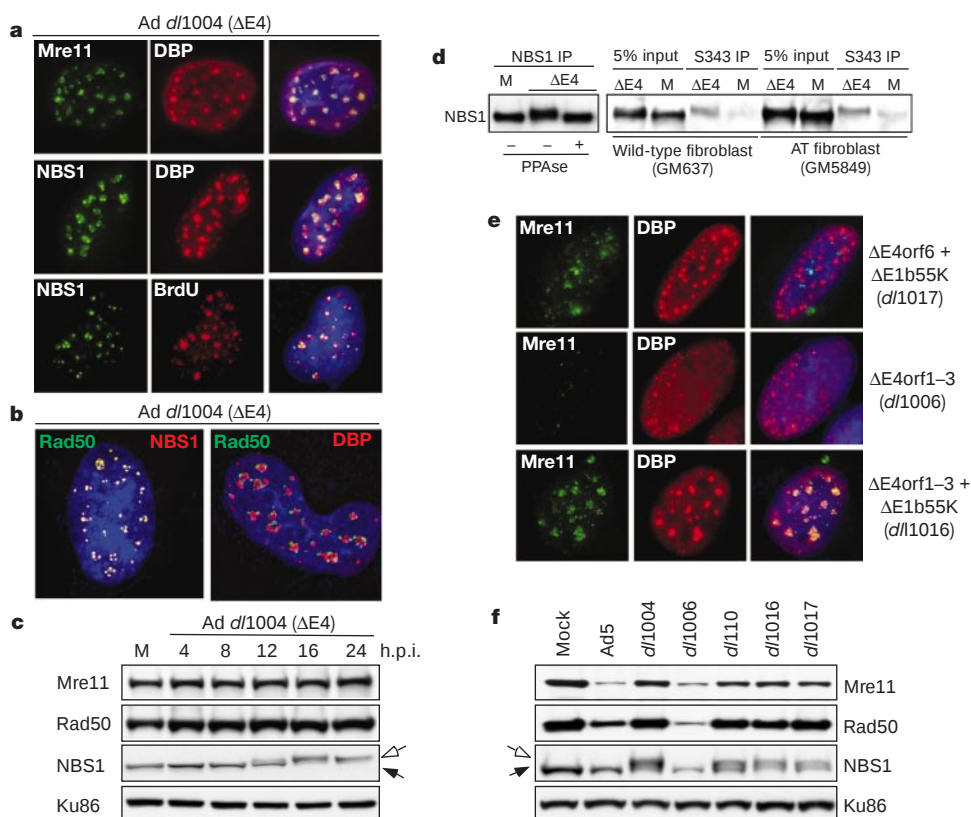


Figure 3 The Mre11–Rad50–NBS1 complex accumulates at viral replication centres in the absence of E4. **a**, HeLa cells were infected with the E4 mutant virus *dl1004* (MOI 25 p.f.u. per cell) and proteins were visualized at 18 h.p.i. by indirect immunofluorescence. Merged images with DAPI staining are shown on the right. Infected cells were pulsed with BrdU for 15 min before fixation, and visualized after double-labelling immunofluorescence with antibodies to BrdU and NBS1. **b**, Images obtained by deconvolution microscopy show that the Rad50 and NBS1 proteins are colocalized in foci (left), which are positioned around the viral replication centres (right). **c**, In the absence of the E4 region, Mre11 and Rad50 proteins are not downregulated and NBS1 is modified. Immunoblotting of lysates from mock-treated (M) or infected HeLa cells at the indicated

times after infection. Open arrow indicates slower migrating forms of NBS1. **d**, The altered migration is due to phosphorylation of NBS1. Upper bands were lost on phosphatase treatment and immunoprecipitated with a phospho-specific antibody. **e**, A combination of E4orf3 and E4orf6–E1b55K prevents accumulation of Mre11–Rad50–NBS1 complexes at viral replication centres. HeLa cells were infected with E4 mutant viruses and analysed by immunofluorescence. **f**, During infections, the E4orf6–E1b55K complex is required for downregulating Mre11, Rad50 and NBS1 proteins and for preventing phosphorylation of NBS1. Cells infected with the indicated viruses were collected at 24 h.p.i. and subjected to immunoblotting. Open arrow indicates slower migrating forms of NBS1.

genome concatemerization, cells expressing wild-type and mutant versions of E4orf3 and E4orf6 were infected with E4-deleted virus *d11004*. Analysis of viral DNA by PFGE showed that wild-type proteins rescued mutant virus from concatemer formation (Fig. 4c, d). By contrast, the E4orf6.AXA mutant was unable to prevent end-joining of viral genomes. We generated a single-point mutation in a conserved region of E4orf3 (E4orf3.N82A) that does not disrupt ND10 and Mre11–Rad50–NBS1 (T.H.S. and M.D.W., unpublished data); this mutant was unable to prevent concatemer formation.

The terminal protein of Ad is a precursor polypeptide (pTP) that is covalently attached to the viral genome and has been implicated in protecting viral termini²³. Immunoblotting showed similar amounts of pTP during infection with mutant viruses (Fig. 5a), which confirmed that concatemer formation was not caused by a decrease in pTP production. The nuclease activity of the Mre11 complex is required to remove covalently attached Spo11 protein from double-strand breaks during meiotic recombination^{24,25}. The Mre11, Rad50 and NBS1 proteins may be part of a complex that also

removes the terminal protein from the Ad genome, generating ends that allow concatemer formation through the non-homologous end-joining pathway. To examine this possibility, we mutated a phosphoesterase signature motif of Mre11 that affects the endonuclease activity of the yeast homologue²⁴ but does not ablate its interaction with Rad50 (ref. 26). Expression of both wild-type and mutant Mre11 (Mre11-3) proteins in ATL3 cells resulted in stabilization of Rad50 and NBS1 (Fig. 5b), as well as relocalization of Rad50 to the nucleus (Fig. 5c). The Mre11-3 protein forms foci upon ionizing irradiation and is recruited to virus centres with the E4-deleted *d11004* (T.H.S. and M.D.W., unpublished data); however, Ad concatemer formation did not occur in Mre11-3-expressing cells infected with *d11004* (Fig. 5d). This suggests that the nuclease activity of the Mre11 complex is likely to be an essential prerequisite to end-joining of viral genomes.

We have shown that the Mre11–Rad50–NBS1 complex can be regulated by degradation and mislocalization. The importance of inactivating this complex by Ad is highlighted by the evolution of redundant strategies to prevent concatemerization of viral DNA. Targeting of Mre11 is also conserved among human Ad serotypes 2, 4, 9 and 12 (T.H.S. and M.D.W., unpublished data). The Mre11–Rad50–NBS1 complex has been studied predominantly by inducing double-stranded breaks with ionizing irradiation^{4,16}. We have shown that virus infection can provide a system for analysing the function of the Mre11–Rad50–NBS1 complex in human cells. This complex has been implicated as a factor that prevents normal cells from malignancy⁴ and is an important regulator of genome stability. Notably, both E4orf3 and E4orf6 proteins transform rodent cells in combination with viral E1 proteins, using a ‘hit-and-run’ mechanism in which viral gene products are undetectable in transformed cells⁷. Our observations support the hypothesis that modulation of DNA repair factors contributes to the ‘hit-and-run’ transformation seen with these and other viral oncogenes. □

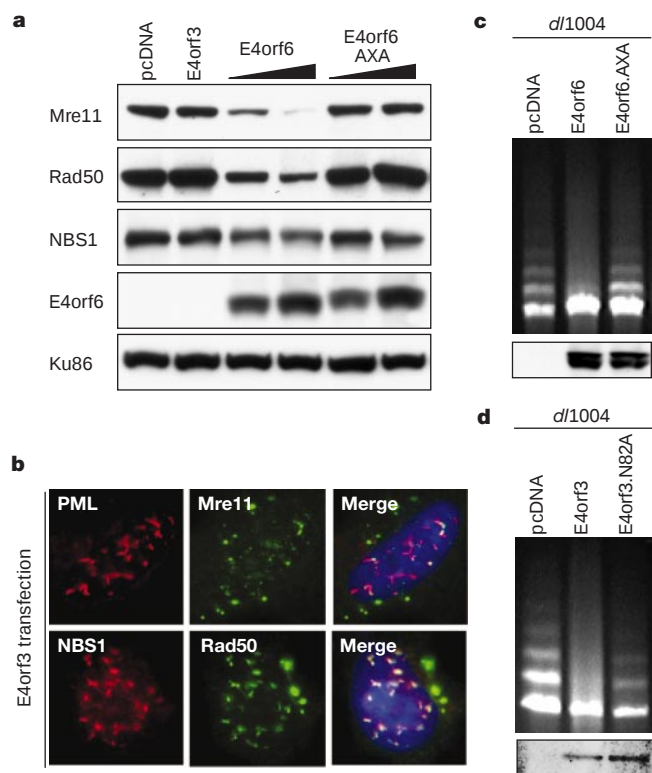


Figure 4 Concatemer formation during Ad infection is prevented by both degradation of cellular DNA repair proteins by the E4orf6–E1b55K complex and redistribution of the Mre11–Rad50–NBS1 complex by E4orf3. **a**, A functional E4orf6–E1b55K complex is required for degradation. Expression vectors were transfected into human 293 cells, and lysates were collected for immunoblotting after 36 h. Neither E4orf3, nor a mutant of E4orf6 that is unable to form a functional complex with the E1b55K protein²⁰, can degrade the cellular proteins. **b**, Expression of E4orf3 alone by transfection results in disruption of the PODs/ND10 and redistribution of Mre11, Rad50 and NBS1 proteins. HeLa cells were transfected and proteins detected by indirect immunofluorescence. Merged images with DAPI staining of the nuclei are shown on the right. **c**, A functional E4orf6–E1b55K complex can prevent concatemer formation. Wild-type and mutant E4orf6 proteins were expressed in HeLa cells by transfection before infection with *d11004*. At 24 h.p.i., cells were collected and their DNA was analysed by PFGE. Shown below is an immunoblot confirming that E4 proteins were expressed. **d**, Wild-type but not mutant E4orf3 protein can rescue E4-deleted viruses from concatemerization. HeLa cells were transfected with expression constructs for E4orf3 or a mutant that cannot disrupt Mre11–Rad50–NBS1 (E4orf3.N82A). Cells were subsequently infected with *d11004* and collected for PFGE and immunoblotting with an antibody to HA to detect tagged E4orf3 proteins.

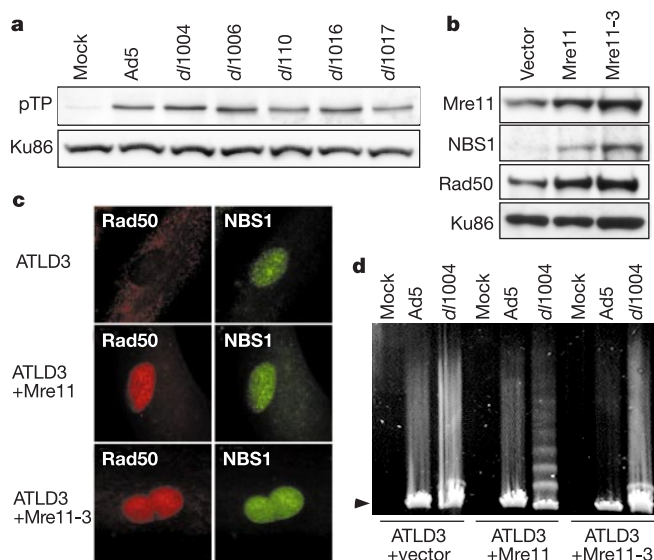


Figure 5 Concatemer formation requires an intact nuclease domain in Mre11. **a**, Amounts of pTP protein are unaffected by mutations in the E4 region. Immunoblotting of lysates from HeLa cells collected at 24 h.p.i. using antisera to pTP and Ku86. **b**, Rad50 and NBS1 proteins are stabilized in ATL3 cells expressing wild-type Mre11 or the mutant Mre11-3. An immunoblot of ATL3 cells transduced with retrovirus vectors is shown. **c**, Wild-type Mre11 and the Mre11-3 mutant both restore nuclear localization of Rad50 in ATL3 cells. Rad50 and NBS1 proteins were detected by indirect immunofluorescence. **d**, Concatemer formation is restored in ATL3 cells expressing wild-type Mre11 but not the Mre11-3 mutant. The position of the single linear viral genome is indicated.

Methods

Cell lines

All cell lines were obtained from the American Tissue Culture Collection or the Coriell Institute apart from ATL1 and ATL3 cells (a gift from J. Petrini), ATM cells (a gift from Y. Shiloh and J. Engelhardt), and cells with mutant ligase IV (180BRM; a gift from H. Wang and G. Iliakis). To complement cells, the complementary DNAs for Mre11 and NBS1 were amplified by PCR and cloned under control of the CMV promoter in a modified version of the pCLNC retrovirus vector²⁷. Retroviruses were generated by transfecting gag/pol packaging cell lines, together with a plasmid expressing the VSV-G envelope protein. ATL1 or NBS1 cells were infected with retrovirus supernatants and selected in 900 µg ml⁻¹ G418. We assessed expression of Mre11 and NBS1 by immunoblotting and used pools of resistant clones for complementation experiments. The *mre11-3* mutant contains the HD129/130LV mutation generated by site-directed mutagenesis using QuickChange (Stratagene).

Antibodies

We visualized viral replication centres by staining for DBP with either a mouse monoclonal antibody (B6-8) or rabbit polyclonal antisera. Antibodies were purchased from Novus (NBS1, ATM), Genetex (Mre11-12D7, Rad50-13B3), Santa Cruz (haemagglutinin A (HA), Ku86, DNA-PKcs, NBS1-S343) and Roche (BrdU). Antisera to E4orf6, pTP, PML, RPA70 and DBP were gifts from P. Branton, J. Schaack, T. Sternsdorf, T. Melendy, A. Levine and P. van der Vliet, respectively. All secondary antibodies were from Jackson Laboratories.

Viruses, infections, PFGE and western blotting

The E4 mutant viruses have been described^{8,11} and were propagated and titred by plaque assays on W162, a Vero-derived E4-complementing cell line²⁸. Wild-type Ad type 5 (Ad5) and dl110 were propagated in human 293 cells. We purified all viruses by two sequential rounds of ultracentrifugation in CsCl gradients and stored them in 40% glycerol at -20°C. HeLa cells were infected with a multiplicity of infection (MOI) of 25 plaque-forming units (p.f.u.) per cell and collected at the indicated hour post-infection (h.p.i.). Other cell lines were infected with MOIs of 25–100 and collected between 48 and 72 h.p.i. We carried out PFGE essentially as described^{5,6} and included λ DNA size markers. Western blotting was done as described²⁰.

Immunofluorescence

Cells were grown on glass coverslips in 24-well dishes and infected with wild-type or mutant viruses at an MOI of 25 p.f.u. per cell. After 8–18 h, the cells were washed with PBS and fixed at -20°C for 10 min with ice-cold methanol:acetone (1:1). We carried out immunofluorescence essentially as described²⁹. In all cases, control staining experiments showed no crossreaction between the fluorophores, and images obtained by staining with individual antibodies were the same as those shown for double-labelling. We stained nuclear DNA with 4',6-diamidino-2-phenylindol (DAPI) and mounted coverslips using Fluoromount-G (Southern Biotechnology Associates) or Vectashield (Vector Labs). Immunoreactivity was visualized by using a Nikon microscope in conjunction with a CCD camera (Cooke Sensicam), or by deconvolution using a DeltaVision microscope system with software from the supplier (Applied Precision). Images were obtained in double or triple excitation mode and processed using SlideBook and Adobe Photoshop.

Received 2 April; accepted 13 May 2002; doi:10.1038/nature00863.

- Haber, J. E. The many interfaces of Mre11. *Cell* **95**, 583–586 (1998).
- Petrini, J. H. The mammalian Mre11–Rad50–nbs1 protein complex: integration of functions in the cellular DNA-damage response. *Am. J. Hum. Genet.* **64**, 1264–1269 (1999).
- Zhu, X. D., Kuster, B., Mann, M., Petrini, J. H. & Lange, T. Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres. *Nature Genet.* **25**, 347–352 (2000).
- Carney, J. P. et al. The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell* **93**, 477–486 (1998).
- Weiden, M. D. & Ginsberg, H. S. Deletion of the E4 region of the genome produces adenovirus DNA concatemers. *Proc. Natl Acad. Sci. USA* **91**, 153–157 (1994).
- Boyer, J., Rohleder, K. & Ketner, G. Adenovirus E4 34k and E4 11k inhibit double strand break repair and are physically associated with the cellular DNA-dependent protein kinase. *Virology* **263**, 307–312 (1999).
- Nevels, M., Tauber, B., Spruss, T., Wolf, H. & Dobner, T. “Hit-and-run” transformation by adenovirus oncogenes. *J. Virol.* **75**, 3089–3094 (2001).
- Bridge, E. & Ketner, G. Redundant control of adenovirus late gene expression by early region 4. *J. Virol.* **63**, 631–638 (1989).
- Halbert, D. N., Cutt, J. R. & Shenk, T. Adenoviral early region 4 encodes functions required for efficient DNA replication, late expression, and host cell shutoff. *J. Virol.* **56**, 250–257 (1985).
- Weinberg, D. H. & Ketner, G. Adenoviral early region 4 is required for efficient viral DNA replication and for late gene expression. *J. Virol.* **57**, 833–838 (1986).
- Bridge, E. & Ketner, G. Interaction of adenoviral E4 and E1b products in late gene expression. *Virology* **174** (1990).
- Riballo, E. et al. Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient. *Curr. Biol.* **9**, 699–702 (1999).
- Lombard, D. B. & Guarente, L. Nijmegen breakage syndrome disease protein and MRE11 at PML nuclear bodies and meiotic telomeres. *Cancer Res.* **60**, 2331–2334 (2000).
- Maser, R. S. et al. Mre11 complex and DNA replication: linkage to E2F and sites of DNA synthesis. *Mol. Cell. Biol.* **21**, 6006–6016 (2001).
- Wu, G., Lee, W. H. & Chen, P. L. NBS1 and TRF1 colocalize at promyelocytic leukemia bodies during late S/G2 phases in immortalized telomerase-negative cells. *J. Biol. Chem.* **275**, 30618–30622 (2000).
- Nelms, B. E., Maser, R. S., MacKay, J. F., Lagally, M. G. & Petrini, J. H. *In situ* visualization of DNA double-strand break repair in human fibroblasts. *Science* **280**, 590–592 (1998).

- Zhao, S. et al. Functional link between ataxia-telangiectasia and Nijmegen breakage syndrome gene products. *Nature* **405**, 473–477 (2000).
- Gatei, M. et al. ATM-dependent phosphorylation of nibrin in response to radiation exposure. *Nature Genet.* **25**, 115–119 (2000).
- Wu, X. et al. ATM phosphorylation of Nijmegen breakage syndrome protein is required in a DNA damage response. *Nature* **405**, 477–482 (2000).
- Cathomen, T. & Weitzman, M. D. A functional complex of adenovirus proteins E1B-55 kDa and E4orf6 is necessary to modulate the expression level of p53 but not its transcriptional activity. *J. Virol.* **74**, 11407–11412 (2000).
- Querido, E. et al. Degradation of p53 by adenovirus E4orf6 and E1B55K proteins occurs via a novel mechanism involving a Cullin-containing complex. *Genes Dev.* **15**, 3104–3117 (2001).
- Doucas, V. et al. Adenovirus replication is coupled with the dynamic properties of the PML nuclear structure. *Genes Dev.* **10**, 196–207 (1996).
- de Vries, E., van Driel, W., Bergsma, W. G., Arnberg, A. C. & van der Vliet, P. C. HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex. *J. Mol. Biol.* **208**, 65–78 (1989).
- Moreau, S., Ferguson, J. R. & Symington, L. S. The nuclease activity of Mre11 is required for meiosis but not for mating type switching, end joining, or telomere maintenance. *Mol. Cell. Biol.* **19**, 556–566 (1999).
- Haber, J. E. A super new twist on the initiation of meiotic recombination. *Cell* **89**, 163–166 (1997).
- Bressan, D. A., Olivares, H. A., Nelms, B. E. & Petrini, J. H. Alteration of N-terminal phosphoesterase signature motifs inactivates *Saccharomyces cerevisiae* Mre11. *Genetics* **150**, 591–600 (1998).
- Naviaux, R. K., Costanzi, E., Haas, M. & Verma, I. M. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. *J. Virol.* **70**, 5701–5705 (1996).
- Weinberg, D. H. & Ketner, G. A cell line that supports the growth of a defective early region 4 deletion mutant of human adenovirus type 2. *Proc. Natl Acad. Sci. USA* **80**, 5383–5386 (1983).
- Weitzman, M. D., Fisher, K. J. & Wilson, J. M. Recruitment of wild-type and recombinant adenovirus-associated virus into adenovirus replication centers. *J. Virol.* **70**, 1845–1854 (1996).

Acknowledgements

We thank G. Ketner for the E4 mutant viruses and W162 cells, I. Verma for retrovirus constructs and cell lines, T. Paull for cDNAs; B. Gilbert for technical assistance; S. Mangel for complemented NBS cells; J. Bailis for advice on PFGE; M. Blower and the Center for Cytometry and Molecular Imaging at the Salk Institute for deconvolution assistance; J. Simon for help with figures; J. Weitzman, M. Grifman, T. Cathomen, C. Barlow, R. Bushman and S. Forsburg for critically reading the manuscript; R. Evans, T. Hunter, G. Wahl, B. Sefton, D. Spector, T. de Lange and J. Petrini for discussions; S. Stampfer for encouragement; and the James B. Pendleton Charitable Trust for providing the Pendleton Microscopy Facility. M.D.W. is supported in part by a grant from NIH, and by gifts from the Joe W. & Dorothy Dorsett Brown Foundation and the Lebensfeld Foundation. T.S. was supported by an NIH Graduate Training Grant to UCSD, and by fellowships from the Chapman Foundation, the Legler Benbough Foundation and the Salk Institute Association.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.D.W. (e-mail: weitzman@salk.edu).

corrigendum

The mammalian sodium channel BNC1 is required for normal touch sensation

Margaret P. Price, Gary R. Lewin, Sabrina L. McIlwraith, Chun Cheng, Jinghui Xie, Paul A. Heppenstall, Cheryl L. Stucky, Anne G. Mannsfeldt, Timothy J. Brennan, Heather A. Drummond, Jing Qiao, Christopher J. Benson, Deirdre E. Tarr, Ron F. Hrstka, Baoli Yang, Roger A. Williamson & Michael J. Walsh

Nature **407**, 1007–1011 (2000).

In this Letter, the x axes for the stimulus response functions shown in Fig. 2a–f were incorrectly labelled 5, 10, 20, 40 (and 80) µm. The axes should read 50, 100, 200, 400 (and 800) µm. This error does not alter our conclusions. □

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)
Amelie Pequignot (4974)
Laura Pearson (4977)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)
Scandinavia: Sille Opstrup (4994)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@nature.com

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.johnson@naturejpn.com

Asia-Pacific Advertising Manager:

Kevyn Johnson

naturejobs

A dual existence

Two London scientists — Ken Powell, chief executive of Arrow Therapeutics, and Peter Shepherd, chairman and chief scientific officer of Xcellsys — were this month honoured for their industrial acumen. But they could equally be recognized for their agility in moving between academia and industry, and not losing contact with either.

Powell, who received the London Biotechnology Network's Entrepreneur of the Year Award (which was co-sponsored by *Naturejobs*), has a history of bridging sectors. As deputy director of the Wolfson Institute for Biomedical Research, he developed the institute's facilities and infrastructure. Before that, as head of biology at the Wellcome Foundation (now part of GlaxoSmithKline), he handled licensing deals. And, in setting up four separate biotech companies, he managed to maintain and expand his academic ties.

Shepherd, who received the London Biotechnology Network's Young Entrepreneur Award, has managed to live in both worlds simultaneously. In addition to his position at Xcellsys, of which he is a co-founder, Shepherd is a professor of biochemistry at University College London. He still carries a full teaching load there and continues to run a lab.

Career paths like his and Powell's were once unusual, Shepherd says. Academics who turned to industry were considered pariahs — once you chose that path, there was no turning back. Now more scientists are discovering that it makes sense to travel back and forth. New ideas come out of academia, but companies are better positioned to turn them into commercial products. And scientists who are comfortable in both spheres can recognize which ideas can become products, and assist in the translation. "It certainly is an advantage to have a foot in both worlds," says Shepherd.

Paul Smaglik

Naturejobs editor



Contents

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS